LHC-NET: BRIDGING LLM PRIORS AND HYPERGRAPH STRUCTURE VIA CHANNEL RETRIEVAL FOR BIOCHEMICAL FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Multivariate Time Series Forecasting (MTSF) serves as a core instrument in AI for Science for analyzing complex biochemical processes . However, practical applications face a fundamental dilemma: balancing the physical deficiency caused by Channel Independence (CI) strategies against the noise overfitting induced by Channel Dependence (CD) strategies. While CI mitigates high-dimensional noise, it severs essential internal physical coupling; conversely, CD attempts to model global dependencies but introduces significant spurious correlations, leading to redundancy and overfitting. To address this, we propose **LHC-Net** (LLM-initialized Hypergraph Channel-Retrieval Network), a novel paradigm integrating scientific domain knowledge with deep learning. Distinct from existing methods that employ Large Language Models (LLMs) as direct numerical predictors, we leverage the general scientific knowledge of LLMs to extract sparse hypergraph structural priors, characterizing high-order mechanistic dependencies among variables. To mitigate the risk of LLM hallucinations, we define this topology as learnable parameter matrices. This allows the model to dynamically correct priors via gradient backpropagation from observational data, achieving a robust integration of mechanistic priors and empirical evidence. Furthermore, we design a temporal channel retrieval strategy that enables the model to access critical variable information from the mechanistic hypergraph on-demand, guided by future temporal contexts. This structured channel fusion strategy restores the necessary physical coupling of dynamical systems while effectively shielding against irrelevant interference. Extensive experiments on a million-scale real-world biopharmaceutical dataset demonstrate that LHC-Net significantly outperforms state-of-the-art models in accuracy. Moreover, it successfully recovers reaction paths consistent with biochemical principles. Crucially, by dynamically grounding LLM priors in observational data, LHC-Net effectively mitigates hallucination risks, significantly enhancing the application potential of AI4S. The code related to this work is available at `https://anonymous.4open.science/r/LHC-Net-8B41`.

1 INTRODUCTION

Biochemical processes are typical complex dynamical systems widely prevalent in pharmaceuticals and synthetic biology Sun et al. (2024); Lin-Gibson & Srinivasan (2019). These systems are governed by complex metabolic networks and physicochemical reactions, involving interactions among multiple state variables Yang & Ge (2021). Accurate prediction of these key indicators is crucial for controlling reaction processes and understanding biological evolution. However, inferring system dynamics and performing Multivariate Time Series Forecasting (MTSF) from observational data remains a significant challenge in AI for Science (AI4S) Ansari et al. (2023). This difficulty arises from the black-box nature of mechanisms and the high-dimensional, non-stationary characteristics of the data.

Despite significant progress in MTSF via deep learning, existing mainstream paradigms exhibit a critical limitation in modeling variable channel relationships. These paradigms generally fall into two static strategies Han et al. (2024): Channel Independence (CI) Nie et al. (2022); Zhang et al. (2022) or Channel Dependence (CD) Zhang & Yan (2022); Wu et al. (2021).. CI strategies effectively avoid noise interference from irrelevant variables by processing each channel independently Chen et al.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

(2024). However, they completely ignore critical coupling relationships in biochemical reactions, such as the causality between bacterial concentration and growth. Conversely, CD strategies model global dependencies through sequence or graph neural networks to achieve channel fusion. Yet, their static, indiscriminative fully-connected design introduces excessive noise, leading to model redundancy and overfitting.

These static split-or-merge strategies fail in biochemical contexts due to the time-varying dynamics of the process. For instance, pharmaceutical fermentation typically evolves through distinct stages: lag and early log phases, log phase, and stationary phase Chater (2006). During the metabolically active log phase, variables show strong coupling Hopwood (2007); Ke et al. (2025). In the stationary phase, however, relationships may decouple or exhibit increased latency. Traditional static models cannot adapt to these dynamic fusion requirements. Furthermore, they overlook a fundamental attribute: biochemical systems are neither isolated variables nor chaotic fully-connected networks, but dynamical systems with highly modular coupling characteristics.

To address this challenge, hypergraph theory provides an ideal mathematical tool for characterizing biochemical modularity Cai et al.. Unlike ordinary graphs that describe pairwise relationships, a hyperedge in a hypergraph can connect an arbitrary number of nodes Wang et al. (2023). This is physically isomorphic to the substrate-enzyme-product synergetic sets in biochemical reactions. However, constructing high-quality biochemical hypergraphs in real-world scientific scenarios faces a severe cold-start problem Duta & Liò (2024). Scientific data is often accompanied by high-dimensional noise and sparsity. Purely data-driven structure learning is computationally difficult and prone to misleading random perturbations, resulting in the learning of numerous spurious correlations Karpatne et al. (2017).

The emergence of Large Language Models (LLMs) offers a solution to break this bottleneck by leveraging their vast general scientific knowledge Zheng et al. (2025). LLMs can infer potential mechanistic associations based on variable semantic names. This inspires a new approach: we can utilize LLMs to construct a semantic-based structural prior. Physically coupled variables are pre-clustered into the same hyperedge, providing the model with a high-quality initialized topology Jin et al. (2024a); Feng et al. (2024).

Nevertheless, in specific scientific domains, LLMs face the challenge of hallucination, potentially providing erroneous priors that contradict specific process realities. Therefore, the core question transforms into: *How to design an architecture that utilizes LLM priors for structured channel fusion while retaining data-driven error correction to prevent misleading priors?* Huang et al. (2025)

Based on this, we propose **LHC-Net** (**L**LM-initialized **H**ypergraph **C**hannel-Retrieval Network). The framework achieves the unification of priors and data through three synergistic components. The *LLM-based Structural Initialization* (**L**SI) leverages domain knowledge to initialize the hypergraph incidence matrix as learnable parameters, allowing observational gradients to fine-tune the topology and correct potential hallucinations. Subsequently, the *Hypergraph Interaction Encoder* (**H**IE) aggregates high-order dependencies via message passing to compensate for the physical deficiency of CI strategies. Finally, the *Channel-Retrieval Decoder* (**C**RD) addresses time-varying dynamics by using future temporal contexts to dynamically retrieve key variables from mechanistic knowledge, realizing denoised channel fusion. The main contributions of this paper are as follows:

1. We propose LHC-Net, which integrates LLM semantic priors with hypergraph neural networks. The LLM warm-start and data fine-tuning paradigm cleverly resolves the trade-off between physical deficiency and noise overfitting.

2. We design a synergistic architecture comprising the HIE and CRD. This achieves a closed-loop prediction from static mechanistic aggregation to dynamic temporal retrieval, precisely adapting to the dynamic channel changes in biochemical contexts.

3. Extensive experiments on a million-scale real-world biopharmaceutical dataset demonstrate that LHC-Net significantly outperforms state-of-the-art models. Visualization confirms that the model successfully recovers key reaction paths consistent with biochemical principles, offering high interpretability. Furthermore, the model exhibits strong robustness against potential hallucination risks, demonstrating significant application value.

2 Related Work and Background

2.1 Related Work

Balancing Channel Independence (CI) and Channel Dependence (CD) is a critical challenge in MTSF Han et al. (2024). CI methods, such as PatchTST Nie et al. (2022), mitigate noise but fail to capture physical couplings, whereas CD methods like Crossformer Zhang & Yan (2022),HMT Song et al. (2024b) and LCCH Song et al. (2024a) often introduce redundancy. Although researchers have employed GNNs to model sparse dependencies, methods like Gelo Sun et al. (2024) are limited to pairwise interactions and cannot capture the high-order modularity inherent in biochemical systems. Hypergraphs theoretically resolve this by connecting arbitrary nodes, yet existing temporal hypergraph approaches rely on data-driven structure learning. In noisy scientific data, these methods suffer from spurious correlations and poor cold-start performance, limiting their effectiveness in capturing true mechanistic topologies.

Incorporating external knowledge offers a solution to the limitations of data-driven structure learning. Current LLM-based paradigms, such as Time-LLM Jin et al. (2024b) and OFA Zhou et al. (2023), predominantly function as direct predictors, generating future sequences via modality alignment. However, in scientific domains, this approach is prone to hallucinations that violate underlying process logic. Distinct from this direct prediction paradigm, LHC-Net utilizes LLMs as structural initializers. We leverage mechanistic knowledge to construct a sparse hypergraph prior while retaining data-driven fine-tuning capabilities. This design resolves the CI-CD conflict by implementing a retrieval-based channel fusion, balancing high-order physical dependencies with robustness against noise.

2.2 Background

Erythromycin fermentation via *Streptomyces erythraeus* serves as a prototypical complex biochemical system. As articulated in Sec. 1, its temporal evolution is governed by dynamic metabolic networks that traverse three distinct physiological phases: the lag and early log phase, the log phase, and the stationary phase. Each phase is characterized by unique modular dependencies and shifting functional logic. Throughout the fermentation lifecycle, the system state is monitored and controlled via a specific set of physical and biochemical variables. To formalize our modeling framework, we denote these critical factors using the following Greek notations:

Substrate Kinetics: Total sugars (σ), non-reducing sugars (κ), and reducing sugars (ρ).

Metabolic Byproducts: Alcohol (γ).

Environmental Controls: External airflow (ψ), dissolved oxygen (ω), and pH (P).

Physical State & Yield: Bacterial density (ϕ), medium viscosity (η), and final chemical potency (α).

The detailed domain knowledge regarding the metabolic switching are thoroughly elaborated in **Appendix B**.

3 Methodology

3.1 Problem Definition

Let $\mathbf{X}_{in} = \{\mathbf{x}_1, \ldots, \mathbf{x}_L\} \in \mathbb{R}^{L \times D}$ denote the input historical observation data sequence, with L representing the lookback window length. At each time step t, the observation vector is defined as $\mathbf{x}_t = [x_t^{k_1}, x_t^{k_2}, \ldots, x_t^{k_D}]^\top \in \mathbb{R}^D$, where each dimension $d \in \{1, \ldots, D\}$ is explicitly indexed by its corresponding semantic descriptor $k_d \in \mathcal{K}$. Here, $\mathcal{K} = \{k_1, k_2, \ldots, k_D\}$ represents the set of domain-specific definitions (e.g., viscosity, potency) that characterize the biochemical identity of each numerical channel. Our objective is to learn a mapping function $\mathcal{F}$ to predict the future sequence $\mathbf{Y} \in \mathbb{R}^{H \times D}$, where H denotes the prediction horizon.

Distinct from purely data-driven methods that treat variables as anonymous channels, we leverage the semantic set $\mathcal{K}$ to inject biochemical priors. We formulate the pre-trained LLM as a frozen knowledge

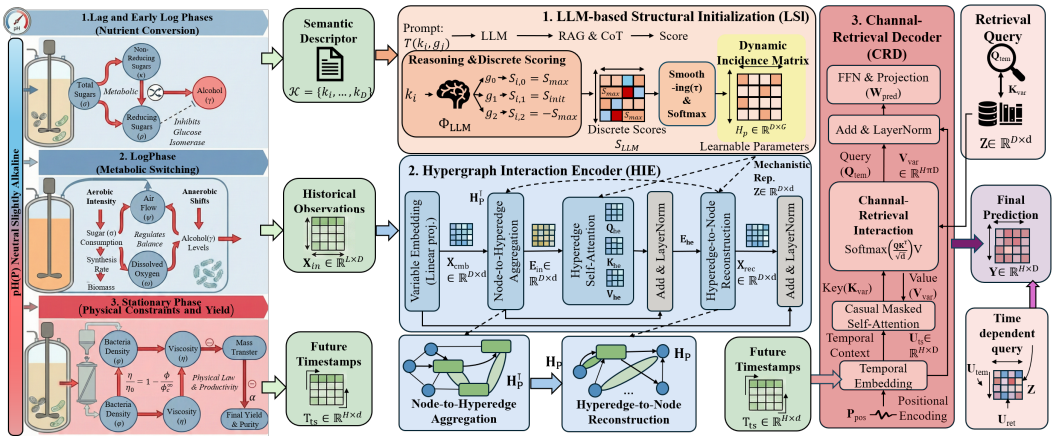

Figure 1: Architecture of LHC-Net for biochemical process modeling.

extraction operator Φ_{LLM} that maps the unstructured text into a structured physical prior $\mathcal{P}$:

$$\mathcal{P} = \Phi_{LLM}(\mathcal{K}) \tag{1}$$

Specifically, $\mathcal{P}$ is instantiated as a hypergraph incidence matrix $\mathbf{H}_p \in \mathbb{R}^{D\times G}$, where G represents the number of latent mechanistic groups. The elements of $\mathbf{H}_p$ quantify the mechanistic affinity between the variable k_d and functional groups. Consequently, the prediction task is reformulated as a conditional learning process based on both data and structural priors:

$$\mathbf{Y} = \mathcal{F}(\mathbf{X}_{in}, \mathbf{H}_p) \tag{2}$$

This formulation ensures that the information aggregation is guided by biochemical principles rather than relying solely on data correlations.

3.2 LLM-BASED STRUCTURAL INITIALIZATION (LSI)

To address the cold-start problem of structure learning in industrial data, the LSI component leverages the domain knowledge of LLMs to provide physical priors for hypergraph construction.

3.2.1 STRUCTURAL PRIOR ELICITATION

To construct a sparse hypergraph topology with explicit physical meaning, we need to quantify the association strength between D process variables and G latent mechanistic groups. Distinct from simple zero-shot prompting, we propose a scoring paradigm based on private knowledge base augmentation Edge et al. (2024) and Chain-of-Thought (CoT) reasoning Wei et al. (2022).

First, we construct a private knowledge base $\mathcal{B}$ incorporating biochemical reaction mechanisms, metabolic pathways, and expert experience. Next, we define a structured prompt template $\mathcal{T}(\cdot)$ to guide LLM reasoning. For the text description k_i of the i-th variable and the definition g_j of the j-th mechanistic group, the prior score generation process is:

$$\mathbf{S}_{ij} = \Phi_{LLM}\left(\text{CoT}(\mathcal{T}(k_i, g_j)) \mid \mathcal{B}\right) \tag{3}$$

To filter noise and obtain a robust sparse structure, we enforce the LLM to output scores within an integer domain. The generated prior score matrix $\mathbf{S}_{LLM}$ satisfies:

$$\mathbf{S}_{LLM} \in \{s \in \mathbb{Z} \mid -S_{max} \le s \le S_{max}\}^{D\times G} \tag{4}$$

In this system, the scores possess explicit topological meanings:

1. $\mathbf{S}_{ij} = S_{max}$ (Core Driver): The variable is a direct participant in the mechanistic group, corresponding to a strong connection.

2. $\mathbf{S}_{ij} \in (0, S_{max})$ (Related Coupling): The variable is significantly influenced by or indirectly associated with the mechanism.

3. $\mathbf{S}_{ij} = -S_{max}$ (Structural Isolation): There is no mechanistic link. This mandatory cut-off mechanism ensures high sparsity of the initialized matrix, preventing erroneous message passing between irrelevant variables.

3.2.2 SMOOTH-SCALED INITIALIZATION

To obtain a smooth initial distribution and avoid gradient saturation Feng et al. (2024), we scale $\mathbf{S}_{LLM}$ by a smoothing coefficient τ to initialize the learnable parameter matrix $\mathbf{H}_{LLM} \in \mathbb{R}^{D\times G}$:

$$\mathbf{H}_{LLM} = \mathbf{S}_{LLM}/\tau \tag{5}$$

During training, we normalize along the hyperedge dimension via the Softmax function to obtain the dynamic hypergraph incidence matrix $\mathbf{H}_p \in \mathbb{R}^{D\times G}$:

$$(\mathbf{H}_p)_{ij} = \frac{\exp((\mathbf{H}_{LLM})_{ij})}{\sum_{k=1}^{G} \exp((\mathbf{H}_{LLM})_{ik})} \tag{6}$$

Here, $\mathbf{H}_p$ retains the physical priors of the LLM while allowing data-driven fine-tuning via backpropagation. To provide a more intuitive understanding of how LLMs and expert knowledge collaborate within LSI, we present a case analysis of LSI in **Appendix K**, utilizing a combination of text and illustrations. Please refer to **Appendix K** and Fig.9 for details. For the contents of the knowledge base, please refer to the **Appendi O**.

3.3 HYPERGRAPH INTERACTION ENCODER (HIE)

The HIE component captures high-order mechanistic dependencies among variables through a variable-hyperedge-variable message passing mechanism. Here, we model the D process variables as hypergraph nodes and the G mechanistic groups as hyperedges.

Distinct from traditional time-slice embedding, we transform the raw input $\mathbf{X}_{in} \in \mathbb{R}^{L\times D}$ into a variable-centric view. We map the temporal dimension to high-dimensional features via a linear layer. Let $\mathbf{W}_{emb} \in \mathbb{R}^{L\times d}$ be the learnable projection matrix. The variable features $\mathbf{X}_{emb} \in \mathbb{R}^{D\times d}$ are obtained as:

$$\mathbf{X}_{emb} = \mathbf{X}_{in}^{\top}\mathbf{W}_{emb} \tag{7}$$

where d denotes the hidden dimension of the model. The encoding process consists of three mathematical steps:

3.3.1 NODE-TO-HYPEREDGE AGGREGATION

Using the incidence matrix $\mathbf{H}_p \in \mathbb{R}^{D\times G}$ generated by LSI as the aggregation operator, we project features from the node space to the mechanistic group space. We calculate the initial hyperedge representation $\mathbf{E}_{in} \in \mathbb{R}^{G\times d}$:

$$\mathbf{E}_{in} = \mathbf{H}_p^{\top}(\mathbf{X}_{emb}\mathbf{W}_{node}) \tag{8}$$

where $\mathbf{W}_{node} \in \mathbb{R}^{d\times d}$ is the node projection matrix. This step aggregates information from variables with similar physical properties into unified functional group representations.

3.3.2 HYPEREDGE-LEVEL INTERACTION

To model system-level interactions between different mechanistic groups, we perform mechanistic self-attention in the hyperedge space. Letting $\mathbf{E}_{in}$ serve as the query, key, and value, we compute the mechanistic projection features:

$$\mathbf{Q}_{he} = \mathbf{E}_{in}\mathbf{W}_{he}^{Q}, \mathbf{K}_{he} = \mathbf{E}_{in}\mathbf{W}_{he}^{K}, \mathbf{V}_{he} = \mathbf{E}_{in}\mathbf{W}_{he}^{V} \tag{9}$$

where $\mathbf{W}_{he}^{Q}, \mathbf{W}_{he}^{K}, \mathbf{W}_{he}^{V} \in \mathbb{R}^{d\times d_k}$ are learnable weight matrices, and d_k is the projection dimension of the attention head. Subsequently, we calculate the attention weight distribution and the updated mechanistic context $\mathbf{E}_{mec}$:

$$\mathbf{A}_{he} = \text{Softmax}\left(\frac{\mathbf{Q}_{he}(\mathbf{K}_{he})^{\top}}{\sqrt{d_k}}\right) \tag{10}$$

$$\mathbf{E}_{mec} = \mathbf{A}_{he}\mathbf{V}_{he} \tag{11}$$

Finally, we obtain the interacted hyperedge features $\mathbf{E}_{he}$ via residual connection and layer normalization:

$$\mathbf{E}_{he} = \text{LayerNorm}\left(\mathbf{E}_{in} + \mathbf{E}_{mec}\mathbf{W}_{he}^{O}\right) \tag{12}$$

3.3.3 HYPEREDGE-TO-NODE RECONSTRUCTION

Finally, we redistribute the optimized mechanistic information back to the variable space. We perform reverse mapping using the incidence matrix $\mathbf{H}_p$ and recover variable features $\mathbf{X}_{rec} \in \mathbb{R}^{D\times d}$ via a reconstruction projection matrix $\mathbf{W}_{rec} \in \mathbb{R}^{d\times d}$:

$$\mathbf{X}_{rec} = (\mathbf{H}_p\mathbf{E}_{he})\mathbf{W}_{rec} \tag{13}$$

The final output $\mathbf{Z} \in \mathbb{R}^{D\times d}$, defined as the global mechanistic knowledge representation for the subsequent CRD component, is obtained via a residual connection:

$$\mathbf{Z} = \text{LayerNorm}(\mathbf{X}_{emb} + \mathbf{X}_{rec}) \tag{14}$$

3.4 CHANNEL-RETRIEVAL DECODER (CRD)

The CRD component transforms the global mechanistic knowledge representation $\mathbf{Z}$ generated by the HIE into future time-series predictions.

3.4.1 INPUT DEFINITION

The CRD component receives two core inputs, representing mechanistic knowledge and temporal context, respectively:

1. Global Mechanistic Knowledge Representation $\mathbf{Z} \in \mathbb{R}^{D\times d}$: The output from the HIE component, serving as an external knowledge base for physical dependencies among variables.

2. Temporal Query Embedding $\mathbf{U}_{ts} \in \mathbb{R}^{H\times d}$: This represents the temporal context for the prediction horizon H. Since future observations are unknown, we construct a pure time representation using known timestamp features (e.g., hour-of-day, day-of-week). Let $\mathbf{T}_{ts} \in \mathbb{R}^{H\times d}$ denote the matrix of these known timestamp features. $\mathbf{U}_{ts}$ is generated via a learnable linear projection augmented with positional information Vaswani et al. (2017):

$$\mathbf{U}_{ts} = \mathbf{T}_{ts}\mathbf{W}_{ts} + \mathbf{P}_{pos} \tag{15}$$

where $\mathbf{W}_{ts} \in \mathbb{R}^{d\times d}$ is the time projection matrix and $\mathbf{P}_{pos} \in \mathbb{R}^{H\times d}$ represents fixed positional encodings. Here, $\mathbf{U}_{ts}$ acts as a temporal context, projecting future time steps into the high-dimensional latent space to guide the subsequent retrieval of mechanistic features.

3.4.2 TEMPORAL SELF-CORRELATION MODELING

First, to capture dependencies between future time steps, we perform causal masked multi-head self-attention. This ensures that the prediction at step t relies only on historical information prior to t, preventing information leakage.

Letting $\mathbf{U}_{ts}$ serve as the source for query, key, and value, we calculate the projection features:

$$\mathbf{Q}_s = \mathbf{U}_{ts}\mathbf{W}_s^{Q}, \quad \mathbf{K}_s = \mathbf{U}_{ts}\mathbf{W}_s^{K}, \quad \mathbf{V}_s = \mathbf{U}_{ts}\mathbf{W}_s^{V} \tag{16}$$

where $\mathbf{W}_s^{Q}, \mathbf{W}_s^{K}, \mathbf{W}_s^{V} \in \mathbb{R}^{d\times d_k}$ are learnable projection matrices. To enforce causal constraints, we introduce a mask matrix $\mathbf{M} \in \mathbb{R}^{H\times H}$, where $\mathbf{M}_{ij} = 0$ if $i > j$, and $\mathbf{M}_{ij} = -\infty$ otherwise. The temporal attention scores and the updated context $\mathbf{U}_s$ are calculated as:

$$\mathbf{A}_s = \text{Softmax}\left(\frac{\mathbf{Q}_s(\mathbf{K}_s)^{\top}}{\sqrt{d_k}} + \mathbf{M}\right) \tag{17}$$

$$\mathbf{U}_s = \mathbf{A}_s\mathbf{V}_s \tag{18}$$

Finally, we obtain the features with temporal context, $\mathbf{U}_{tem}$, via a residual connection and layer normalization:

$$\mathbf{U}_{tem} = \text{LayerNorm}\left(\mathbf{U}_{ts} + \mathbf{U}_s\mathbf{W}_s^{O}\right) \tag{19}$$

where $\mathbf{W}_s^{O} \in \mathbb{R}^{d\times d}$ is the output projection matrix.

3.4.3 CHANNEL RETRIEVAL INTERACTION

After establishing the temporal context, the model executes the core channel retrieval process to extract key information from the mechanistic knowledge base $\mathbf{Z}$ generated by the encoder. We treat the temporal features $\mathbf{U}_{tem}$ as the query, and the mechanistic representation $\mathbf{Z}$ as both key and value:

$$\mathbf{Q}_{tem} = \mathbf{U}_{tem}\mathbf{W}_{cr}^{Q}, \quad \mathbf{K}_{var} = \mathbf{Z}\mathbf{W}_{cr}^{K}, \quad \mathbf{V}_{var} = \mathbf{Z}\mathbf{W}_{cr}^{V} \tag{20}$$

$$\mathbf{A}_{ret} = \text{Softmax}\left(\frac{\mathbf{Q}_{tem}(\mathbf{K}_{var})^{\top}}{\sqrt{d}}\right) \tag{21}$$

$$\mathbf{U}_{ret} = \mathbf{A}_{ret}\mathbf{V}_{var} \tag{22}$$

where $\mathbf{W}_{cr}^{Q}, \mathbf{W}_{cr}^{K}, \mathbf{W}_{cr}^{V}$ are projection matrices. $\mathbf{U}_{ret} \in \mathbb{R}^{H\times d}$ represents the mechanistic features dynamically retrieved and aggregated from the variable knowledge base according to the requirements of the current time step.

3.4.4 FUSION AND GENERATION

Finally, we fuse the retrieved mechanistic information into the temporal flow and generate the final prediction via a feed-forward network (FFN):

$$\mathbf{U}_{out} = \text{LayerNorm}\left(\mathbf{U}_{tem} + \mathbf{U}_{ret}\right) \tag{23}$$

$$\mathbf{Y} = \text{FFN}(\mathbf{U}_{out})\mathbf{W}_{pred} \tag{24}$$

where $\mathbf{W}_{pred} \in \mathbb{R}^{d\times D}$ is learnable projection matrices, which projects $\mathbf{U}_{out}$ onto the variable dimension.The final output $\mathbf{Y} \in \mathbb{R}^{H\times D}$ represents the multivariate time series prediction result.

4 EXPERIMENT

4.1 DATA ACQUISITION & PHASE SEGMENTATION

We utilize the Erythromycin Pharmaceutical Dataset (EPD), a large-scale proprietary dataset collected from a real-world biopharmaceutical production line throughout 2024. Sampled at a 30-second frequency, the dataset comprises 1,048,576 timestamps. To comprehensively evaluate LHC-Net against non-stationary biochemical dynamics, we segment the dataset into three sub-datasets based on fermentation kinetics: lag and early log phases (Sub1), log phase (Sub2), and stationary limiting phase (Sub3). Standard real-world data preprocessing is applied to each sub-dataset, including the removal of missing values and statistical outlier rejection. We adopt a strict chronological splitting strategy to prevent information leakage, dividing each sub-dataset into training, validation, and testing sets in a 7:2:1 ratio.

Notably, identical Greek symbols across sub-datasets denote data from the same physical sensors (e.g., viscosity η in Sub1 and Sub3). While the physical units remain consistent, their biochemical roles evolve across different time windows. This design rigorously tests the model's ability to capture functional shifts of the same variable under different kinetic constraints. Table 2 and Fig. 3 in **Appendix C.1** present the statistical characteristics of representative variables, exhibiting highly heterogeneous patterns ranging from Gaussian to right-skewed and bimodal distributions.

Furthermore, to strongly validate the broad applicability and generalization capabilities of LHC-Net beyond the proprietary EPD, we incorporate IndPenSim, an internationally recognized, large-scale benchmark for penicillin fermentation simulation. To demonstrate the model's versatility across diverse operational scenarios, we select two distinct sub-tasks from this benchmark: *Exhaust Gas Analysis* and *Vessel Control*. Detailed descriptions of this public dataset, including specific variable selections and preprocessing procedures, are provided in **Appendix C.2**.

4.2 EXPERIMENTAL SETTINGS

Addressing the specific biochemical contradictions in each phase, we incorporate expert knowledge to categorize variables into $G = 3$ latent functional modules. Table 3 in **Appendix D** details the evolution

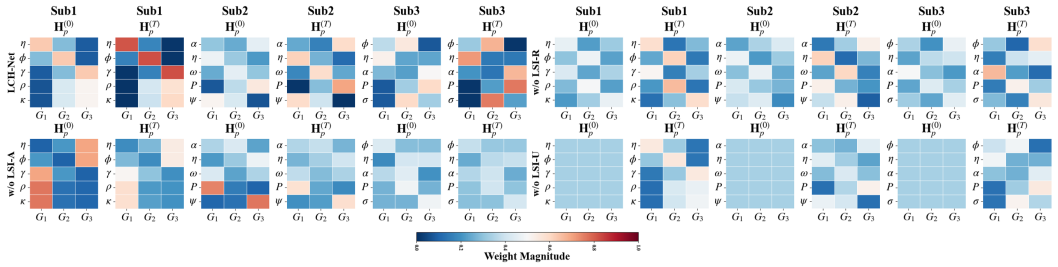

Figure 2: Evolution of hypergraph topological weights and anti-hallucination analysis

of these physical definitions. All experiments are conducted on 8 × NVIDIA RTX 4090 GPUs using PyTorch supported by 256GiB of system memory. We employ MSE and MAE as evaluation metrics. Training utilizes the Adam optimizer (initial learning rate 10^{-4}, cosine decay, batch size 512) with early stopping. For consistency, we set $L = 96$ and $H \in \{96, 192, 336, 720\}$. LHC-Net configurations include Gemini-3-pro Gemini Team (2025) knowledge extraction ($S_{max} = 5$), $\tau = 5$, $d = 256$, $d_k = 64$, and 0.2 dropout. Regarding the rationale for selecting Gemini-3-pro as the knowledge extractor, we also compare the performance and inference times of nine different LLMs in **Appendix J**. Among these, τ and L are the key hyperparameters of LHC-Net. The results of our sensitivity analysis for these two parameters can be found in **Appendices L.1 and L.2**. Moreover we used two layers of HIE and one layer of CRD. Twelve state-of-the-art (SOTA) baselines across three paradigms are implemented following their optimal reported settings. The details of these baselines are shown in the Table 4 in the **Appendix D**

4.3 Main Results

Table 1 presents the average performance of LHC-Net and competitive baselines across all prediction horizons ($H \in \{96, 192, 336, 720\}$). Detailed experimental results for each specific horizon are provided in **Appendix E**. LHC-Net achieves state-of-the-art performance across all proprietary EPD and public IndPenSim sub-datasets. By leveraging LLM-initialized hypergraphs for structured channel fusion, LHC-Net effectively overcomes the physical decoupling of Channel-Independent (CI) methods (e.g., PatchTST) and the noise-overfitting of Channel-Dependent (CD) strategies (e.g., Crossformer). Furthermore, it significantly outperforms static graph models (e.g., HMT) and direct LLM predictors like Time-LLM (reducing MSE by ∼30% on Sub2), proving that injecting LLMs as structural priors is optimal for capturing complex biochemical dynamics. Additionally, its superior metrics on the public IndPenSim tasks (Exhaust Gas Analysis and Vessel Control) strongly confirm its broad generalizability. A comprehensive performance analysis and detailed baseline comparisons are provided in **Appendix E**.

Moreover, detailed visualizations demonstrating how the Channel Retrieval Decoder (CRD) dynamically characterizes stage-specific biochemical processes to transcend conventional CI/CD baselines are presented in **Appendix H**. LHC-Net not only demonstrates exceptional predictive capabilities but also outperforms other LLM-based models in terms of overall training time and potential for industrial application. Tests regarding LLM-related applications can be found in **Appendix J**, while a comprehensive analysis of computational efficiency is provided in **Appendix M**. Statistical significance tests comparing LHC-Net against the second-best performer, HMT, are detailed in **Appendix F**. Additionally, ablation studies regarding the contributions of core components (LSI, HIE, CRD) are provided in **Appendix G**, and further visualization analysis elucidating the superior performance of LHC-Net from a data distribution perspective is presented in **Appendix I**.

4.4 Analysis of Anti-Hallucination Capabilities

The risk of LLM hallucination in scientific domains is a primary concern for AI4S. To validate LHC-Net's ability to rectify erroneous priors via data-driven learning, we visualize the evolution of the hypergraph incidence matrix from initialization $\mathbf{H}_p^{(0)}$ to the converged state $\mathbf{H}_p^{(T)}$ (Fig. 2). We examine three simulated hallucination scenarios defined in **Appendix G**: ignorant (w/o LSI-U), noisy (w/o LSI-R), and counter-factual (w/o LSI-A).

Table 1: Average performance comparison across all sub-datasets. The **best** and second-best results for each sub-dataset are highlighted.

Model	Sub1		Sub2		Sub3		Exhaust Gas		Vessel Control	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Crossformer	0.494	0.448	6.156	0.678	6.318	0.620	0.952	0.451	0.749	0.477
HMT	0.363	0.377	4.417	0.582	4.306	0.520	0.978	0.432	0.833	0.515
PatchTST	0.502	0.453	6.253	0.654	6.251	0.627	1.531	0.639	1.296	0.506
FEDformer	0.487	0.468	6.630	0.716	6.520	0.652	1.835	0.666	0.914	0.817
DLinear	0.484	0.482	5.197	0.698	5.092	0.622	1.261	0.575	0.867	0.575
CEGlo	0.940	0.714	6.918	0.873	6.850	0.846	2.018	0.851	1.331	0.726
LightTS	0.410	0.429	5.381	0.761	5.279	0.669	1.141	0.520	0.783	0.518
LCCH	0.465	0.455	4.517	0.641	4.363	0.567	1.107	0.533	0.843	0.560
Time-LLM	0.452	0.422	6.085	0.668	5.974	0.595	2.345	0.538	0.952	0.503
iTransformer	0.492	0.443	5.900	0.670	6.295	0.619	1.926	0.614	0.883	0.520
DHSL	0.365	0.380	4.332	0.577	4.266	0.513	0.993	0.474	0.734	0.490
TimeMKG	0.453	0.418	5.883	0.653	6.148	0.589	1.677	0.635	0.978	0.580
LHC-Net	**0.347**	**0.368**	**4.249**	**0.571**	**4.167**	**0.513**	**0.951**	**0.416**	**0.706**	**0.453**

1. Data-Driven Error Correction: The w/o LSI-A variant provides the most compelling evidence of robustness. Despite being initialized with adversarial structures (e.g., severing critical biomass-substrate links), its heatmap undergoes drastic reconstruction during training, with the weight distribution converging towards the morphological pattern of standard LHC-Net. This demonstrates that the HIE effectively utilizes gradients from ground-truth observations to override erroneous priors. Similarly, w/o LSI-U and w/o LSI-R evolve from uniform blur or random noise into distinct modular contours consistent with LHC-Net. This confirms that the mechanistic structure inferred by the LLM aligns with the statistical regularities of the underlying data. However, the failure of w/o LSI-A to fully recover the optimal topology compared to the standard LHC-Net underscores the necessity of a high-quality cold start for escaping local optima.

2. Physical Consistency Evolution: Comparing $\mathbf{H}_p^{(T)}$ across sub-datasets reveals that the model captures phase-dependent functional shifts, aligning with the mechanisms described in Sec.2.2:

Sub1 (lag and early log phases): Viscosity η exhibits clear modular separation from metabolic variables. This conforms to the lag and early log phases reality where η acts merely as a physical background, not yet strongly coupled with metabolic flux. **Sub3 (stationary phase)**: A significant shift occurs where η shows high co-occurrence with density ϕ and total sugar σ. This accurately reflects the emergence of non-Newtonian fluid dynamics induced by high biomass density, where viscosity becomes the dominant constraint limiting substrate mass transfer. In summary, LHC-Net employs a dialectical mechanism that grounds LLM priors with data-driven dynamics. This approach effectively resists hallucinations, evolving hypergraph structures that faithfully reflect time-varying physical facts rather than blindly adhering to static knowledge.

5 Conclusion

This paper resolves the trade-off between Channel Independence (CI) physical deficiency and Channel Dependence (CD) noise overfitting in MTSF. We propose LHC-Net, a structured channel retrieval framework for modular, time-varying biochemical processes. By leveraging LLMs to build sparse hypergraph priors, LHC-Net restores high-order physical coupling via message passing, while dynamic temporal retrieval ensures denoised information access. Experiments on biopharmaceutical data confirm that LHC-Net outperforms state-of-the-art baselines and recovers physically consistent reaction paths. Ultimately, integrating LLM structural guidance with data-driven correction yields high-precision, mechanistically faithful models for AI4S. Future work will explore fully automated frameworks minimizing expert intervention; preliminary automated latent group discovery experiments are detailed in **Appendix N**. For detailed information regarding the limitations of LHC-Net, please refer to **Appendix P**.

AI USE STATEMENT

The core intellectual contributions of this paper, including its concepts, methodology, and results, are the original work of the authors. In preparing the paper, we utilized Google's Gemini model for language assistance. Its use was strictly confined to improving grammar, refining prose for clarity, and assisting with the beautification of the typesetting across the entire document. We confirm that no scientific content, text, figures, or code was generated by the AI. All scientific insights are solely the product of the authors' collective efforts.

REPRODUCIBILITY STATEMENT

The full implementation of LHC-Net, including model architecture, training scripts, prompt templates, and RAG pipeline configurations, is released anonymously at `https://anonymous.4open.science/r/LHC-Net-8B41`. The primary EPD dataset is subject to an NDA and cannot be fully released; our industrial partner has authorised a companion dataset of identical sensor configurations at `https://zenodo.org/records/14619074` for verification purposes. The public IndPenSim benchmark is available at `http://www.industrialpenicillinsimulation.com`. All hyperparameters, baseline settings, sensitivity analyses, and statistical significance tests are documented in the appendices.

ETHICS STATEMENT

The EPD dataset was collected under a formal industrial collaboration agreement as part of routine process monitoring; no human subjects or personally identifiable information are involved. The publicly released companion dataset has been approved by our industrial partner with all commercially sensitive identifiers removed. The LLMs in this work are used solely as offline structural prior extractors and are not involved in any safety-critical decisions; hallucination risks are explicitly mitigated via data-driven fine-tuning and validated through ablation studies. We recommend domain-expert validation of LLM-generated priors before production deployment. By reducing model size to 15.9M parameters ($\approx 37\times$ fewer than Time-LLM), LHC-Net also substantially lowers the computational cost and carbon footprint of AI-assisted biopharmaceutical research.

REFERENCES

Hani Nasser Abdelhamid. A review on hydrogen generation from the hydrolysis of sodium borohydride. *international journal of hydrogen energy*, 46(1):726–765, 2021.

Abdul Fatir Ansari, Alvin Heng, Andre Lim, and Harold Soh. Neural continuous-discrete state space models for irregularly-sampled time series. In *International Conference on Machine Learning*, pp. 926–951. PMLR, 2023.

Tingyi Cai, Yunliang Jiang, Ming Li, Lu Bai, Changqin Huang, and Yi Wang. Hypernear: Unnoticeable node injection attacks on hypergraph neural networks. In *Forty-second International Conference on Machine Learning*.

Keith F Chater. Streptomyces inside-out: a new perspective on the bacteria that provide us with antibiotics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1469): 761–768, 2006.

Jialin Chen, Jan Eric Lenssen, Aosong Feng, Weihua Hu, Matthias Fey, Leandros Tassiulas, Jure Leskovec, and Rex Ying. From similarity to superiority: Channel clustering for time series forecasting. *Advances in Neural Information Processing Systems*, 37:130635–130663, 2024.

Jacob Cohen. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin*, 70(4):213, 1968.

Gheorghe Comanici, Eric Bieber, Mike Schaekermann, Ice Pasupat, Noveen Sachdeva, Inderjit Dhillon, Marcel Blistein, Ori Ram, Dan Zhang, Evan Rosen, et al. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *arXiv preprint arXiv:2507.06261*, 2025.

DeepSeek-AI et al. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437*, 2024.

DeepSeek-AI et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025a.

DeepSeek-AI et al. Deepseek-v3.2: Pushing the frontier of open large language models. *arXiv preprint arXiv:2512.02556*, 2025b.

Iulia Duta and Pietro Liò. Sphinx: Structural prediction using hypergraph inference network. *arXiv preprint arXiv:2410.03208*, 2024.

Darren Edge, Ha Trinh, Newman Cheng, Joshua Bradley, Alex Chao, Apurva Mody, Steven Truitt, Dasha Metropolitansky, Robert Osazuwa Ness, and Jonathan Larson. From local to global: A graph rag approach to query-focused summarization. *arXiv preprint arXiv:2404.16130*, 2024.

Yifan Feng, Chengwu Yang, Xingliang Hou, Shaoyi Du, Shihui Ying, Zongze Wu, and Yue Gao. Beyond graphs: Can large language models comprehend hypergraphs? *arXiv preprint arXiv:2410.10083*, 2024.

Google Gemini Team. Gemini 3 pro: Model card and technical documentation. Technical report, Google DeepMind, 2025. URL `https://storage.googleapis.com/deepmind-media/Model-Cards/Gemini-3-Pro-Model-Card.pdf`.

Lu Han, Han-Jia Ye, and De-Chuan Zhan. The capacity and robustness trade-off: Revisiting the channel independent strategy for multivariate time series forecasting. *IEEE Transactions on Knowledge and Data Engineering*, 36(11):7129–7142, 2024.

David A Hopwood. *Streptomyces in nature and medicine: the antibiotic makers*. Oxford University Press, 2007.

Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, et al. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, 43(2):1–55, 2025.

Kalervo Järvelin and Jaana Kekäläinen. Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems (TOIS)*, 20(4):422–446, 2002.

Bowen Jin, Gang Liu, Chi Han, Meng Jiang, Heng Ji, and Jiawei Han. Large language models on graphs: A comprehensive survey. *IEEE Transactions on Knowledge and Data Engineering*, 2024a.

Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-fang Li, Shirui Pan, et al. Time-llm: Time series forecasting by reprogramming large language models. In *International Conference on Learning Representations*, 2024b.

Anuj Karpatne, Gowtham Atluri, James H Faghmous, Marc Steinbach, Ayush Khandelwal, Ricardo Bolt, Shashi Shekhar, Steven V Peng, and Vipin Kumar. Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on Knowledge and Data Engineering*, 29(10):2318–2331, 2017.

Xiang Ke, Shuohan Wang, Xing Jiang, Xu Li, Ju Chu, and Xiwei Tian. Enhancing erythromycin production by saccharopolyspora erythraea e3:: sucba in chemically defined medium. *Bioresource Technology*, pp. 133730, 2025.

Matej Kristan, Aleš Leonardis, and Danijel Skočaj. Multivariate online kernel density estimation with gaussian kernels. *Pattern recognition*, 44(10-11):2630–2642, 2011.

Ajay Kumar and Vishal Gulati. Experimental investigation and optimization of surface roughness in negative incremental forming. *Measurement*, 131:419–430, 2019.

Ganhui Lan and Yuhai Tu. Information processing in bacteria: memory, computation, and statistical physics: a key issues review. *Reports on Progress in Physics*, 79(5):052601, 2016.

Xiaobo Li, Jun Chen, Joakim M Andersen, Ju Chu, and Peter R Jensen. Cofactor engineering redirects secondary metabolism and enhances erythromycin production in saccharopolyspora erythraea. *ACS Synthetic Biology*, 9(3):655–670, 2020.

Sheng Lin-Gibson and Vijay Srinivasan. Recent industrial roadmaps to enable smart manufacturing of biopharmaceuticals. *IEEE Transactions on Automation Science and Engineering*, 18(1):176–183, 2019.

Allison J Lopatkin and James J Collins. Predictive biology: modelling, understanding and harnessing microbial complexity. *Nature Reviews Microbiology*, 18(9):507–520, 2020.

Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. *arXiv preprint arXiv:2211.14730*, 2022.

OpenAI. Gpt-4o system card. *OpenAI Technical Report*, 2024. URL `https://openai.com/index/gpt-4o-system-card/`.

OpenAI. Introducing gpt-5.2. *OpenAI Blog/Technical Report*, 2025a. URL `https://openai.com/index/introducing-gpt-5-2/`.

OpenAI. Gpt-5 system card. *OpenAI Technical Report*, 2025b. URL `https://openai.com/index/gpt-5-system-card/`.

Philip Sedgwick. Pearson's correlation coefficient. *Bmj*, 345, 2012.

Yichen Song, Chenxi Xia, Simengxu Qiao, Changdi Li, Qunshan He, and Xinggao Liu. Lcch: A low computational complexity hybrid model based on the half-router attention for biopharmaceutical indicators prediction. In *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1661–1665. IEEE, 2024a.

Yichen Song, Hu Xu, Changdi Li, Qunshan He, Zijian Tian, and Xinggao Liu. Hmt: Hybrid mechanistic transformer for bio-fabrication prediction under complex environmental conditions. *Expert Systems with Applications*, pp. 124389, 2024b.

Ruoxi Sun, Hanjun Dai, and Adams Wei Yu. Does gnn pretraining help molecular representation? *Advances in Neural Information Processing Systems*, 35:12096–12109, 2022.

Yifei Sun, Junming Liu, Yirong Chen, Xuefeng Yan, and Ding Wang. Timemkg: Knowledge-infused causal reasoning for multivariate time series modeling. *arXiv preprint arXiv:2508.09630*, 2025.

Ziyue Sun, Yinlong Li, Qunshan He, Hu Xu, Wenhai Wang, and Xinggao Liu. Causality enhanced global-local graph neural network for bioprocess factor forecasting. *IEEE Transactions on Industrial Informatics*, 20(10):12428–12438, 2024.

Gemini Team et al. Gemini 2.0 flash: Technical report and model cards. *Google DeepMind Technical Report*, 2024.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Shun Wang, Yong Zhang, Xuanqi Lin, Yongli Hu, Qingming Huang, and Baocai Yin. Dynamic hypergraph structure learning for multivariate time series forecasting. *IEEE Transactions on Big Data*, 10(4):556–567, 2024.

Yuxin Wang, Quan Gan, Xipeng Qiu, Xuanjing Huang, and David Wipf. From hypergraph energy functions to hypergraph neural networks. In *International Conference on Machine Learning*, pp. 35605–35623. PMLR, 2023.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in Neural Information Processing Systems*, 34:22419–22430, 2021.

Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, Xiaojun Chang, and Chengqi Zhang. Connecting the dots: Multivariate time series forecasting with graph neural networks. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 753–763, 2020.

Zeyu Yang and Zhiqiang Ge. Rethinking the value of just-in-time learning in the era of industrial big data. *IEEE Transactions on Industrial Informatics*, 18(2):976–985, 2021.

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pp. 11121–11128, 2023.

Tianping Zhang, Yizhuo Zhang, Wei Cao, Jiang Bian, Xiaohan Yi, Shun Zheng, and Jian Li. Less is more: Fast multivariate time series forecasting with light sampling-oriented mlp structures. *arXiv preprint arXiv:2207.01186*, 2022.

Yunhao Zhang and Junchi Yan. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In *The Eleventh International Conference on Learning Representations*, 2022.

Yizhen Zheng, Huan Yee Koh, Jiaxin Ju, Anh TN Nguyen, Lauren T May, Geoffrey I Webb, and Shirui Pan. Large language models for scientific discovery in molecular property prediction. *Nature Machine Intelligence*, pp. 1–11, 2025.

Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *International Conference on Machine Learning*, pp. 27268–27286. PMLR, 2022.

Tian Zhou, Peisong Niu, Liang Sun, Rong Jin, et al. One fits all: Power general time series analysis by pretrained lm. *Advances in neural information processing systems*, 36:43322–43355, 2023.

A Appendix

B Detailed Biochemical Background of Erythromycin Fermentation

Erythromycin fermentation via *Streptomyces erythraeus* serves as a prototypical complex biochemical system. As articulated in Sec.1, its temporal evolution is governed by dynamic metabolic networks that traverse three distinct physiological phases, each characterized by unique modular dependencies.

1. Lag and Early Log Phase: The initial stage focuses on spore activation and the enzymatic hydrolysis of total sugars (σ), defined as the sum of non-reducing (κ) and reducing sugars (ρ). A critical inhibitory coupling exists within this nutrient assimilation module: elevated levels of alcohol (γ), a byproduct of anaerobic metabolism, feedback to inhibit glucose isomerase, thereby decelerating the conversion from κ to ρ Abdelhamid (2021); Lopatkin & Collins (2020). Accurately capturing this mechanism is prerequisite for modeling the system's start-up trajectory.

2. Log Phase: The system enters its peak productivity era during the metabolically intensive the log phase. Dynamics here are defined by metabolic switching: rapid depletion of substrates (σ) reflects the aerobic metabolic drive, while alcohol fluctuations (γ) signal anaerobic shifts. This balance is strictly regulated by external airflow (ψ), which maintains the mass transfer environment for dissolved oxygen (ω). Deviations in ψ or ω can trigger metabolic drift, directly suppressing the biosynthesis rate.

3. Stationary Phase : In the stationary phase, the dominant regime shifts from metabolic reaction to physical limitation. The structural dependence between bacterial density (ϕ) and medium viscosity (η) adheres to a rigorous physical law: $\frac{\eta}{\eta_0} = 1 - \frac{\phi}{\phi_c^\infty}$ Kumar & Gulati (2019). Managing this rheological constraint is paramount, as excessive viscosity impedes mass transfer and downstream filtration, severely compromising the final yield and potency (α) of erythromycin Lan & Tu (2016).

Throughout the lifecycle, pH (P) acts as a global modulator governing bacterial physiology and enzymatic kinetics. A neutral to slightly alkaline environment is essential to maximize chemical potency (α) and overall system throughput Li et al. (2020). These stage-specific mechanisms synthesize diverse factors, encompassing substrate kinetics (σ, ρ, κ), metabolic byproducts (γ), environmental controls (ψ, ω, P), and physical state variables (ϕ, η, α).

C Details of Datasets

C.1 Detailed EPD Description

1. Statistical Characteristics Table 2 presents the descriptive statistics for the core biochemical variables in the EPD dataset. A key observation is the multi-scale nature of the data: variables span multiple orders of magnitude, ranging from pH levels ($P \approx 7.24$) to chemical potency ($\alpha \approx 6588$) and airflow ($\psi \approx 4449$). Additionally, the high standard deviations relative to the means (e.g., viscosity η has a mean of 102.59 with a std dev of 76.43) indicate significant volatility within the fermentation process, necessitating robust normalization and modeling strategies.

2. Distributional Heterogeneity Fig. 3 visualizes the probability density functions of these variables, revealing that real-world biochemical data strictly deviates from ideal Gaussian assumptions. Instead, we observe distinct distributional heterogeneity driven by underlying physical mechanisms. Variables such as dissolved oxygen (ω) and pH (P) exhibit sharp, narrow peaks (tight Gaussian), reflecting the active intervention of PID controllers maintaining setpoints, where minor fluctuations still encode critical state information. In contrast, viscosity (η) and reducing Sugar (ρ) display pronounced right-skewed, heavy-tailed distributions, corresponding to the irreversible accumulation of byproducts or rapid substrate consumption during specific phases. Notably, alcohol level (γ) and total Sugar (σ) manifest distinct bimodal or multimodal distributions. This statistical feature serves as a physical signature of the system switching between different metabolic regimes, providing strong empirical justification for the phase-dependent modeling approach adopted by LHC-Net.

3. Data Transparency Due to Non-Disclosure Agreement (NDA) restrictions imposed by our industrial partner, the full EPD dataset cannot be publicly released in its entirety. However, to

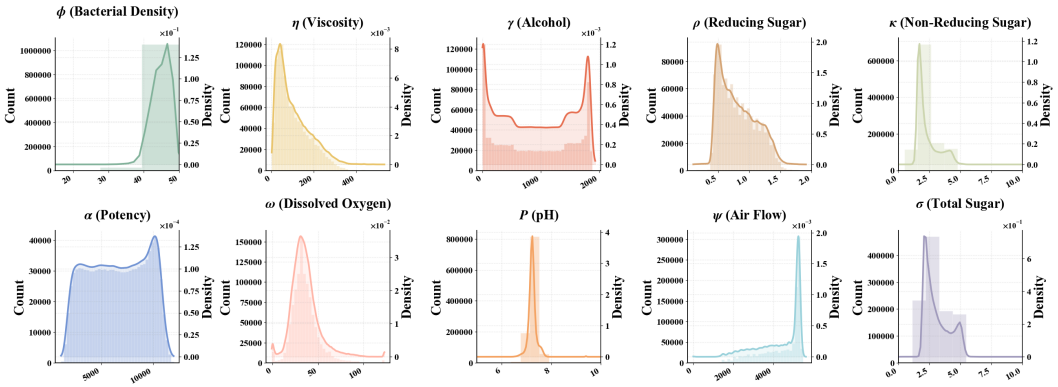

Figure 3: Variable distribution of the EPD

Table 2: Descriptive Statistics of Biochemical Variables

Description	Var.	Mean	Median	Std Dev	25th Per.	75th Per.
Bacterial Density	ϕ	45.02	45.00	7.61	44.00	46.00
Viscosity	η	102.59	80.00	76.43	42.00	148.00
Alcohol Level	γ	902.57	892.00	637.49	295.00	1517.00
Reducing Sugar	ρ	0.81	0.76	0.29	0.56	1.03
Non Reducing Sugar	κ	2.29	1.92	0.97	1.67	2.65
Chemical Potency	α	6588.32	6631.00	2744.68	4206.00	9047.00
Dissolved Oxygen	ω	35.37	33.00	14.89	26.00	42.00
pH Level	P	7.24	7.23	0.40	7.16	7.30
Air Flow	ψ	4449.35	5041.00	1062.22	3790.00	5270.00
Total Sugar	σ	3.11	2.75	1.10	2.25	3.83

ensure reproducibility and to allow the broader research community to verify the statistical properties reported above, our industrial partner has publicly released a companion dataset originating from the same factory and exhibiting identical variable characteristics (i.e., the same set of biochemical sensors, comparable process dynamics, and consistent distributional patterns) (`https://zenodo.org/records/14619074`). This publicly accessible dataset serves as a faithful proxy for validating the descriptive statistics (Table 2) and distributional heterogeneity (Fig. 3) discussed in EPD, thereby substantiating the credibility and generalizability of our empirical findings despite the confidentiality constraints on the original data.

C.2 Detailed IndPenSim Dataset Description

To thoroughly assess the generalization potential of LHC-Net in non-core fermentation processes, we introduced the IndPenSim dataset alongside our proprietary EPD. IndPenSim is a high-fidelity mathematical simulation dataset that closely emulates real-world industrial environments, containing a comprehensive set of 2,239 variables that cover exhaust gas analysis, fermentation processes, vessel control, and Raman spectroscopy.

To complement the core fermentation phase analysis performed on the EPD, we intentionally avoided duplicating similar fermentation phase data and instead formulated two specific sub-tasks:

1. Exhaust Gas Analysis: Focuses on predicting critical metabolic and respiratory indicators, including pH, Carbon Dioxide emission (CO_2), Oxygen Uptake Rate (OUR), Oxygen concentration (O_2), and Carbon Dioxide Evolution Rate (CER).

2. Vessel Control: Aims to forecast the physical and thermodynamic parameters of the reactor, comprising feed flow rate, heating/cooling water flow rate, vessel weight, vessel volume, and generated heat.

Table 3: Evolution of variable grouping logic across fermentation phases based on expert biochemical knowledge.

sub-dataset	Var.	Group	Definition
Sub1 (lag and early log Phases)	$\eta, \phi, \gamma, \rho, \kappa$	g_0	Physical Rheology Baseline
		g_1	Nutrient Assimilation
		g_2	Carbon Metabolism Flow
Sub2 (Log Phase)	$\alpha, \eta, \omega, P, \psi$	g_0	Mass Transfer Environment
		g_1	Aerobic Metabolic Drive
		g_2	Biosynthesis Regulation
Sub3 (Stationary Phase)	$\phi, \eta, \alpha, P, \sigma$	g_0	Rheological Constraint
		g_1	Maintenance Metabolism
		g_2	Product Accumulation

Because IndPenSim consists of simulation data, it inherently avoids the missing values and sensor anomalies typical of raw industrial data. Consequently, the preprocessing pipeline for this dataset was strictly limited to scenario-specific variable extraction and chronological splitting into training, validation, and testing sets in a 7:2:1 ratio. No further data-level modifications, such as imputation or artifact rejection, were applied, thereby ensuring an objective and rigorous evaluation of the model's predictive performance. The data access URL is http://www.industrialpenicillinsimulation.com.

D Detailed Experimental Settings

1. Evolution of Variable Grouping Logic: To capture the phase-dependent metabolic characteristics, the LSI component utilizes a dynamic grouping strategy ($G = 3$) guided by expert knowledge. As detailed in Table 3, the definition of these latent functional modules evolves across sub-datasets. For instance, the grouping logic shifts from prioritizing nutrient assimilation in the lag and early log phases (Sub1) to emphasizing rheological constraints in the final stationary phase (Sub3), ensuring that the initialized hypergraph structure remains physically consistent with the dominant biochemical mechanisms of each stage.

2. Baseline Implementations: We conduct a comprehensive comparison against 10 state-of-the-art baselines categorized into three distinct paradigms: General MTSF (covering both Channel-Independent and Channel-Dependent strategies), Domain-Informed methods (integrating mechanism or causal graphs), and LLM-Based approaches. Table 4 summarizes the classification and core strategies of these competitive models. All baselines are implemented following their official configurations to ensure a fair comparison.

E Complete Main Results under Different Horizons

Table 5 and Table 6 provide a comprehensive breakdown of performance across all prediction horizons ($H \in \{96, 192, 336, 720\}$). Complementing the average results in the main text, these metrics highlight three critical advantages of LHC-Net:

1. Universal Dominance: LHC-Net demonstrates absolute superiority by securing the best MSE performance in 13 out of 15 aggregated settings, as shown in the "WINS" column. This confirms that its performance is consistent across varying temporal scales, not an artifact of averaging.

2. Long-Term Robustness: The performance gap notably widens at extended horizons ($H = 720$). While baselines like iTransformer suffer from overfitting in Sub3, LHC-Net maintains a low error rate. It also consistently outperforms the second-best HMT, proving that dynamic retrieval handles long-term biochemical shifts better than static graph convolutions.

3. Comparison of Structural Prior and Direct Inference: The substantial margin over Time-LLM validates our architectural choice. In high-dimensional biochemical domains, constraining LLM

Table 4: Summary of Baselines and LHC-Net.

Model	Paradigm	Strategy
PatchTST Nie et al. (2022)	**General MTSF**	Channel Independence (CI)
DLinear Zeng et al. (2023)		
LightTS Zhang et al. (2022)		
Crossformer Zhang & Yan (2022)		Channel Dependence (CD)
FEDformer Zhou et al. (2022)		
DHSL Wang et al. (2024)		
HMT Song et al. (2024b)	**Domain-Informed**	Mechanism
LCCH Song et al. (2024a)		
CEGlo Sun et al. (2024)		Causal Graph
Time-LLM Jin et al. (2024b)	**LLM-based**	Reprogramming
TimeMKG Sun et al. (2025)		Knowledge Graph
LHC-Net	**Ours**	**Hypergraph & LLM**

knowledge within a learnable hypergraph topology ($\mathbf{H}_p$) is significantly more effective and robust than using LLMs as direct predictors, which are prone to generative noise.

F Significance Test

To further evaluate model robustness under random perturbations, we conducted statistical significance tests comparing LHC-Net with HMT, the best-performing baseline in the main experiments. With the random seed fixed at 2021, we executed three independent runs for each model across all prediction horizons. Table 7 reports the 95% confidence intervals for MSE and MAE. Based on independent two-sample t-tests, LHC-Net demonstrates statistically significant superiority over HMT in the vast majority of settings ($p < 0.05$), validating the statistical significance of the improvements.

G Ablation Study

G.1 Ablation Study Variants Setting

To verify the necessity of the core components in LHC-Net (LSI, HIE, and CRD), we conducted a comprehensive ablation study. We constructed eight variants across three categories and evaluated their average performance across all sub-datasets under horizons $H \in \{96, 192, 336, 720\}$. Here are three categories:

1. Initialization Variants (w/o LSI): To investigate the quality of priors.

- w/o LSI-U replaces LLM initialization with uniform scores.
- w/o LSI-R uses random noise for hyperedge weights.
- w/o LSI-A employs an adversarial initialization strategy by reversing the sign of LLM-generated scores according to the Eq.equation 3 ($S_{i,j} \to -S_{i,j}$). This constructs a counter-factual error prior, forcing the model to treat strongly correlated variables as irrelevant and vice versa, to test robustness against erroneous knowledge.

2. Encoder Variants (w/o HIE): To investigate topological structures.

- w/o HIE-L removes the hypergraph structure, degenerating into linear embedding.
- w/o HIE-P constructs the graph using static pearson correlation coefficients Sedgwick (2012), whose adjacency matrices are visualized in Fig.5. To ensure the validity of the comparison,

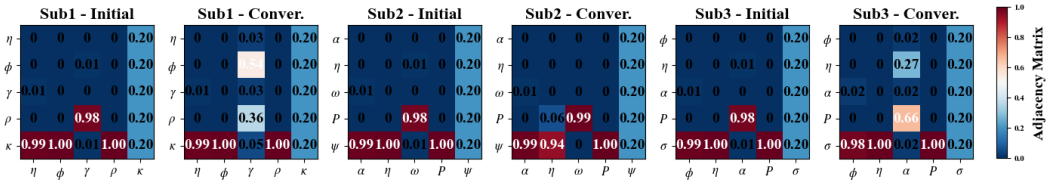

Figure 4: Evolutionary trajectory of adjacency matrices in the w/o HIE-G variant, where 'Conver.' serves as an abbreviation for the converged state

.

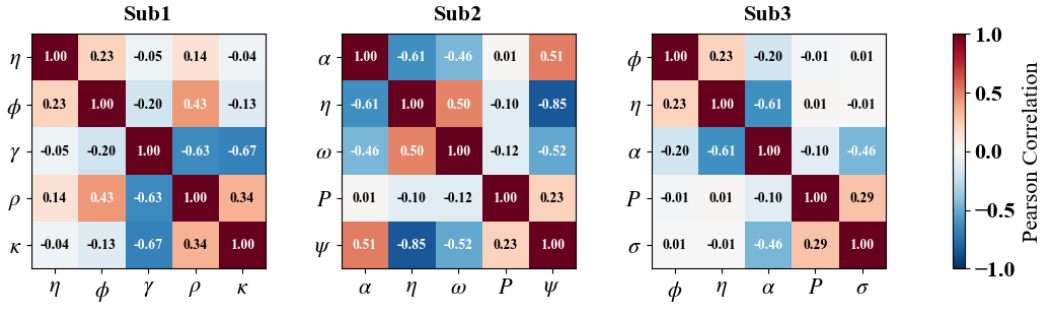

Figure 5: Visualization of the static pearson correlation coefficients utilized in the w/o HIE-P variant.

we pre-computed these coefficients exclusively on the training set and froze them, explicitly excluding them from subsequent parameter updates.

- w/o HIE-G employs a data-driven GNN Sun et al. (2022) for structure learning. To ensure a rigorously fair comparison, the w/o HIE-G variant adopts a graph generation formulation identical to the core HIE component. The critical distinction lies in the source of the priors: whereas HIE leverages LLM-derived knowledge, this variant employs standard random initialization. Distinct from w/o LSI-R which applies static random noise, this variant utilizes randomly initialized yet learnable embeddings to infer latent dependencies from scratch, a standard paradigm validated in graph structure learning Wu et al. (2020).

  Let $\mathbf{E}_1, \mathbf{E}_2 \sim \mathcal{N}(0, 1)$ be the randomly initialized node embeddings. We define the random structural prior as $\mathbf{H}_{rand} = \text{ReLU}(\mathbf{E}_1\mathbf{E}_2^\top)$. The final adjacency matrix $\mathbf{A}_\mathbf{G}$ is normalized as follows, which is similar with the Eq.equation 6:

$$(\mathbf{A}_\mathbf{G})_{ij} = \frac{\exp((\mathbf{H}_{rand})_{ij})}{\sum_{k=1}^{D} \exp((\mathbf{H}_{rand})_{ik})} \tag{25}$$

  We explicitly visualize the evolutionary trajectory of $\mathbf{A}_\mathbf{G}$-from its random initialization to the converged state in Fig. 4.

3. Decoder Variants (w/o CRD): To investigate retrieval and temporal decoding.

- w/o CRD-R removes the retrieval process, severing the link between variables and the knowledge represent.
- w/o CRD-T replaces the temporal decoder with a direct projection output.

G.2 Ablation Study Result

As shown in Fig. 6, our analysis yields three key observations:

Compared to w/o LSI-U and w/o LSI-R, simple structured initialization yields performance degradation. This proves that LHC-Net's gain stems from high-quality LLM priors rather than increased parameter count. Most notably, w/o LSI-A suffers the most severe performance collapse. This demonstrates the directional correctness of LLM-extracted knowledge; enforcing a counter-factual

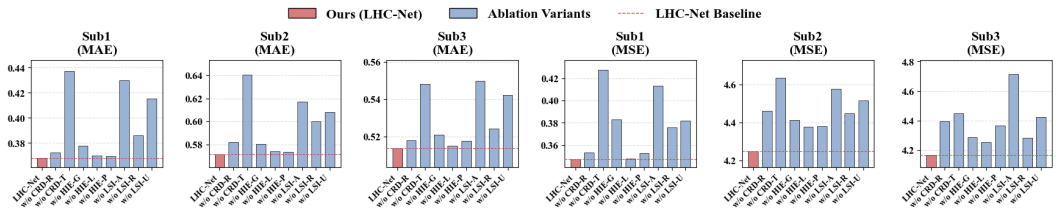

Figure 6: Ablation results for all sub-datasets. Red bars and dashed lines denote the full LHC-Net performance.

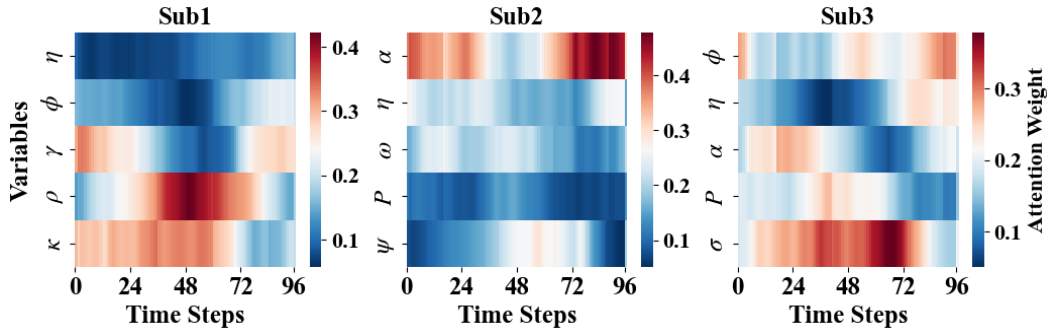

Figure 7: Visualization of dynamic channel retrieval in CRD.

prior that violates physical facts causes greater misleading effects than random noise, inversely validating the accuracy of the LSI component.

w/o HIE-G performs significantly worse than LHC-Net and even underperforms the linear baseline w/o HIE-L on Sub2. As visualized in Fig. 4, the baseline exhibits pathological mode collapse, manifesting as vertical dominance where attention fixates on a single random node. This observation debunks the common assumption that random initialization offers a neutral starting point. Instead, it exposes that random initialization is blindly biased.

The mechanism involves a destructive interaction between activation functions and stochastic noise: ReLU arbitrarily severs $\sim$50% of connections, while Softmax exponentially amplifies minor stochastic variances in the remaining noise.

This "Winner-Takes-All" dynamic causes the model to superstitiously fixate on a random variable solely due to its marginally higher initial variance. This arbitrary preference traps the model in a severe local optimum from the onset-a structural blind spot determined by stochastic luck rather than physical logic, which data-driven gradients struggle to correct.

Additionally, while w/o HIE-P outperforms unstructured models, it lags behind LHC-Net, indicating that linear statistical correlations cannot capture complex high-order non-linear interactions.

The performance drop in w/o CRD-R across all sub-datasets proves that the retrieval mechanism effectively bridges the mechanistic and temporal spaces. Meanwhile, the inferiority of w/o CRD-T indicates that simple direct projection fails to adapt to time-varying characteristics, such as the kinetic drift from the log phase to the stationary phase.

In conclusion, the superiority of LHC-Net stems from the organic integration of semantic initialization, hypergraph modeling, and dynamic knowledge retrieval, with each component being indispensable.

H Analysis of Channel Retrieval

We visualize the retrieval attention map $\mathbf{A}_{ret}$ ($L = H = 96$) in Fig. 7 to verify kinetic consistency. The heatmaps reveal that LHC-Net adopts a dynamic attention pattern, distinctively varying across

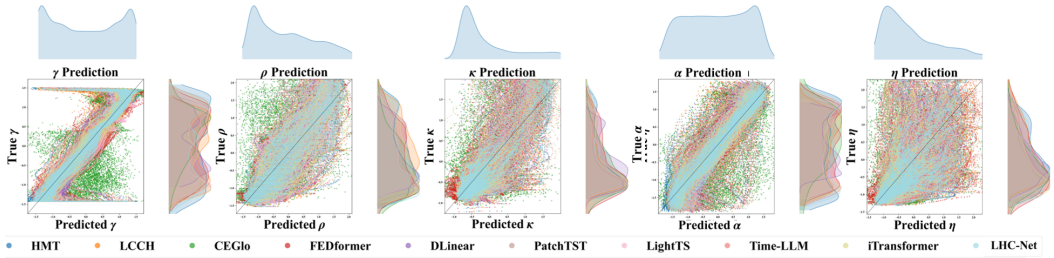

Figure 8: Visualization of prediction performance ($L = H = 96$).

phases. Information regarding the biochemical background of the different phases during the fermentation process can be found in **Appendix B**.

1. Sub1 (Nutrient Assimilation): High attention weights concentrate on non-reducing sugar κ during the early prediction steps. This aligns with the lag and early log phase mechanism, correctly identifying κ as the primary fuel source initiating the biomass growth trajectory.

2. Sub2 (Metabolic Switching): The Fig. 7 displays an alternating high-intensity pattern between airflow ψ and potency α, reflecting their causal input-output coupling during peak metabolism. Conversely, attention on viscosity η fades, showing the model's ability to filter secondary factors to distill the core metabolic flux.

3. Sub3 (Stationary Phase): A sharp increase in attention intensity for total sugar σ is observed in the late horizon (time steps > 80). This indicates the model learns to monitor residual substrate levels to forecast potential depletion risks and process termination.

These visualizations validate LHC-Net's core philosophy: LSI establishes a denoised structural prior, while CRD enables dynamic, demand-driven retrieval. This synergy between static mechanistic injection and adaptive channel awareness achieves interpretability and generalization levels unattainable by conventional CI/CD strategy.

I Visualization of Prediction Performance

To visually evaluate the prediction performance of LHC-Net, we select five representative variables based on the biochemical phase segmentation in Sec.2.2: γ, ρ, κ (core metabolites in nutrient assimilation), α (key product in biosynthesis), and η (complex rheological constraint in the stationary phase). We plot prediction scatter plots and Gaussian Kernel Density Estimation (KDE) Kristan et al. (2011) curves with $L = H = 96$ (Fig. 8).

In the scatter plots, LHC-Net exhibits the tightest convergence toward the identity line $y = x$, significantly outperforming baselines like Time-LLM (pink), which displays notable dispersion on γ and η. The KDE curves further confirm that LHC-Net accurately captures diverse distributional characteristics. Whether for the approximately normal distribution of α or the significantly long-tailed and skewed η, the predicted distribution of LHC-Net aligns closely with the ground truth (the light blue shading above the Fig.8). This indicates that the model captures not only temporal trends but also the complex stochastic statistical laws governing biochemical variables.

J LSI Robustness and Feasibility Testing

To evaluate the robustness and feasibility of the LSI framework, we conduct experiments across nine LLMs representing the current frontier of AI research. These include Google's reasoning-optimized Gemini-3-Pro Gemini Team (2025), the multimodal Gemini-2.5-Pro Comanici et al. (2025), and the high-efficiency Gemini-2-Flash Team et al. (2024). Additionally, we utilized OpenAI's most recent iterations, GPT-5.2 OpenAI (2025a) and GPT-5 OpenAI (2025b), alongside the multimodal flagship GPT-4o OpenAI (2024). To represent specialized architectures, we incorporate the reasoning-incentivized DeepSeek-R1 DeepSeek-AI et al. (2025a) and the high-performance Mixture-of-Experts

(MoE) models DeepSeek-V3.2 DeepSeek-AI et al. (2025b) and DeepSeek-V3 DeepSeek-AI et al. (2024).

We conducted five independent runs per sub-dataset (temperature=0.01), reporting the mode for scores and average for inference time.

The complete experimental results are shown in Table 8.

To rigorously quantify the alignment between the LLM-generated structural priors and expert domain knowledge, we employ two complementary metrics: Quadratic Weighted Kappa (QWK) Cohen (1968) to evaluate the consistency of discrete scoring, and Normalized Discounted Cumulative Gain (NDCG) Järvelin & Kekäläinen (2002) to assess the ranking accuracy of variable importance. The specific definitions are as follows:

1. Quadratic Weighted Kappa (QWK): This metric measures the inter-rater agreement between the LLM-generated scores $\mathbf{S}_{LLM}$ and the expert annotations $\mathbf{S}_{GT}$, penalizing disagreements based on their magnitude. To compute this, we first flatten the scoring matrices into $N = D \times G$ paired samples. We then construct an observed confusion matrix $\mathbf{O}$, where each entry $\mathbf{O}_{uv}$ counts the number of samples assigned score u by the expert and score v by the LLM. $\mathbf{O}_{uv}$ is defined as:

$$\mathbf{O}_{uv} = \sum_{n=1}^{N} \mathbb{I}(\mathbf{S}_{GT}^{(n)} = u \wedge \mathbf{S}_{LLM}^{(n)} = v) \tag{26}$$

Where $\mathbb{I}$ is an indicator function, which evaluates to 1 if the condition is met, and 0 otherwise. Thus, the QWK is defined as:

$$\text{QWK} = 1 - \frac{\sum_{u,v} \mathbf{W}_{uv}\mathbf{O}_{uv}}{\sum_{u,v} \mathbf{W}_{uv}\mathbf{E}_{uv}} \tag{27}$$

Here, the summation indices u and v iterate over all possible discrete score categories. $\mathbf{W}_{uv}$ is the quadratic weight matrix representing the penalty for disagreement:

$$\mathbf{W}_{uv} = \frac{(u-v)^2}{(2S_{max})^2} \tag{28}$$

where $2S_{max}$ represents the maximum possible score difference. $\mathbf{E}_{uv}$ is the expected matrix calculated from the outer product of the marginal histograms, representing the probability of agreement by chance:

$$\mathbf{E}_{uv} = \frac{(\sum_k \mathbf{O}_{uk}) \times (\sum_k \mathbf{O}_{kv})}{N} \tag{29}$$

QWK $\in [-1, 1]$. A value approaching **1** indicates perfect agreement, meaning the LLM precisely captures the exact intensity of mechanistic couplings defined by experts. A value near **0** (or negative) implies that the LLM's scoring consistency is no better than random guessing, failing to reflect valid physical priors.

2. Normalized Discounted Cumulative Gain (NDCG): To evaluate whether the LLM correctly prioritizes core drivers within each functional group, we treat the variable ranking problem as an information retrieval task. NDCG processes the data **column-wise**. For each functional group $j \in \{1, \ldots, G\}$, we calculate the ratio between the model's gain and the ideal gain, then report the mean:

$$\text{NDCG} = \frac{1}{G} \sum_{j=1}^{G} \left(\frac{\text{DCG}_D^{(j)}}{\text{IDCG}_D^{(j)}} \right) \tag{30}$$

The calculation for a specific group j involves four rigorous steps:

Step 1: Relevance Assignment. We define the true utility of the i-th variable in group j using the expert annotation matrix: $rel_i = \mathbf{S}_{GT}[i, j]$.

Step 2: Predicted DCG Calculation. We determine the ranking order based on the LLM's predictions. Let $\pi_j(k)$ be the index of the variable ranked at the k-th position according to the descending order of the LLM scores $\mathbf{S}_{LLM}[:, j]$. The DCG accumulates the expert utility (rel) based on this predicted order:

$$\text{DCG}_D^{(j)} = \sum_{k=1}^{D} \frac{2^{rel_{\pi_j(k)}} - 1}{\log_2(k+1)} \tag{31}$$

Step 3: Ideal DCG (IDCG) Calculation. To normalize the metric, we determine the optimal ranking order. Let $\pi_j^*(k)$ be the index of the variable ranked at the k-th position according to the descending order of the expert scores $\mathbf{S}_{GT}[:, j]$. IDCG represents the maximum possible gain achievable:

$$\text{IDCG}_D^{(j)} = \sum_{k=1}^{D} \frac{2^{rel_{\pi_j^*(k)}} - 1}{\log_2(k+1)} \tag{32}$$

Step 4: Normalization. The final score is the ratio DCG/IDCG, ensuring NDCG $\in [0, 1]$.

NDCG $\in [0, 1]$. A value approaching **1** signifies that the LLM's ranking (π_j) is identical to the expert's ideal ranking (π_j^*) for the top-weighted variables, successfully prioritizing the core drivers of the functional module.

Experimental results shown in Tabel 8 indicate that while expert tacit knowledge is difficult to perfectly replicate, frontier models like Gemini-3-pro and GPT-5.2 exhibit minimal deviation from human judgments, confirming the high semantic penetrability of our prompt template. Crucially, combined with the anti-hallucination findings in Sec.4.4, LHC-Net demonstrates strong error correction capabilities; even models with higher MAE (e.g., Gemini-2-flash) provide sufficient warm starts that are rectified by data-driven fine-tuning. Regarding efficiency, most models complete scoring within 3–6 seconds. Even reasoning-heavy models like DeepSeek-R1 remain within acceptable latency limits for long-cycle fermentation processes, confirming that the LSI component is not a bottleneck for online deployment.

K LLM Reasoning Data Flow and Expert Alignment

To intuitively illustrate the operational mechanism of our proposed hybrid paradigm and to provide transparency into the hypergraph construction process, we present a detailed case study focusing on the fermentation log phase (Sub2). As depicted in Fig. 9, this case visualizes the end-to-end data flow comparing frontline expert annotations with the automated reasoning results generated by the large language model.

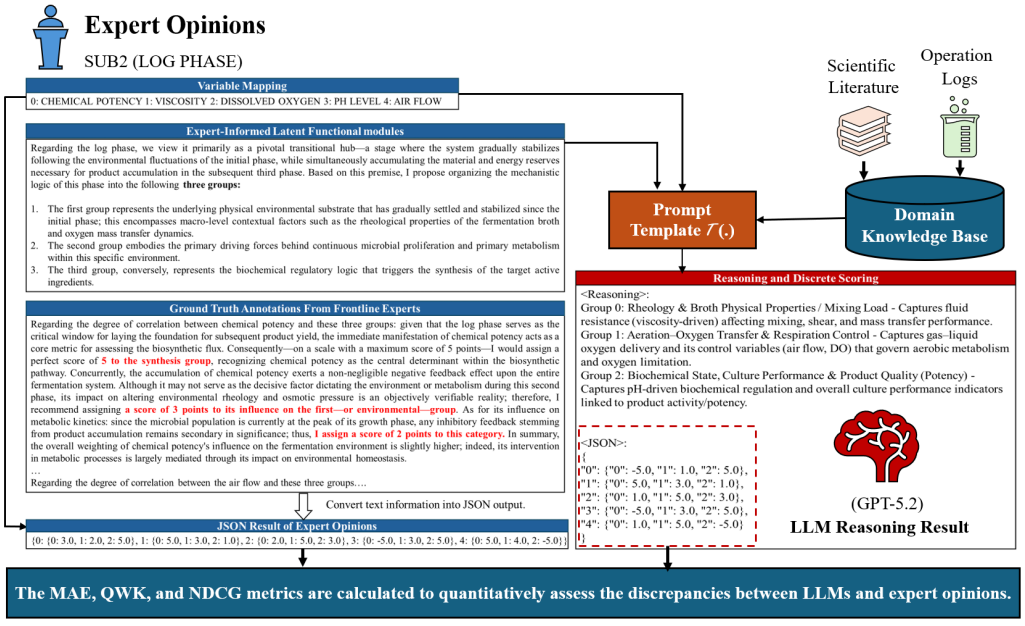

Figure 9: Data flow and alignment analysis between frontline expert opinions and LLM reasoning results for hypergraph construction in the Sub2 (Log Phase) scenario.

The workflow fundamentally begins with the variable mapping, where core physical and biochemical sensors are identified. Following our hybrid design philosophy, frontline experts define the macro-

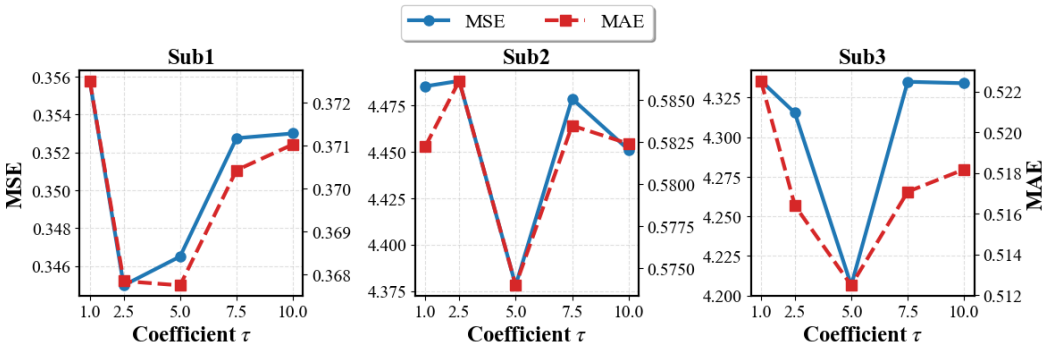

Figure 10: Sensitivity analysis of the smooth coefficient τ across sub-datasets.

level latent functional modules. For the Log Phase, the system is partitioned into three biologically meaningful groups: (1) the physical environmental substrate (e.g., rheology and mass transfer); (2) the driving forces for microbial proliferation and primary metabolism; and (3) the biochemical regulatory logic governing active ingredient synthesis.

On the left side of Fig. 9, the **Ground Truth Annotations** from domain experts are detailed. Experts utilize their experiential knowledge to assign discrete correlation scores between specific variables and the latent groups. For instance, recognizing Chemical Potency as the central determinant of the biosynthetic pathway, experts assign a maximum score of 5 to the synthesis group. Its secondary feedback effect on environmental rheology and metabolic kinetics receives scores of 3 and 2, respectively. These textual rationales are strictly parsed into a structured JSON format to serve as the benchmark.

Concurrently, on the right side of the workflow, the **LLM Reasoning Pipeline** is executed. A Prompt Template $\mathcal{T}(\cdot)$ is formulated, integrating the predefined variable mappings, the expert-informed latent modules, and contextual information retrieved from the customized Domain Knowledge Base (compiled from scientific literature and factory operation logs). The LLM processes this prompt to produce step-by-step reasoning. It accurately maps the variables to the physical, aeration, and biochemical states, ultimately generating its own discrete scoring matrix in JSON format.

Finally, to rigorously and quantitatively assess the discrepancy between the LLM's zero-shot topological reasoning and the human expert's ground truth, the framework calculates comprehensive alignment metrics, including MAE, QWK, and NDCG. This case study not only demonstrates the high interpretability of LHC-Net's internal decision-making process but also validates the efficacy of utilizing LLMs to mine micro-hypergraph associations under the guidance of expert-defined macro-parameters.

L Parameter Sensitivity Analysis

L.1 Impact of Smooth Coefficient

We focus on the LLM score smoothing coefficient τ (Eq. 5), while the analysis of lookback length L is detailed in **Appendix L.2**. Fixing $L = 96$, we evaluate average performance across all horizons. As shown in Fig. 10, a distinct U-shaped trend is observed, achieving optimality at $\tau = 5.0$:

1. Small τ ($\tau < 5.0$) causes over-sharpening. The distribution approaches a one-hot vector, discarding secondary physical connections and destabilizing backpropagation.

2. Large τ ($\tau > 5.0$) leads to over-smoothing. High entropy dilutes specific LLM priors into uniformity, causing the model to lose mechanistic guidance and revert to a cold-start state.

3. Optimal τ ($\tau = 5.0$) balances structural salience with optimization plasticity. It preserves hierarchical priors while maintaining sufficient gradient flow for effective data-driven fine-tuning.

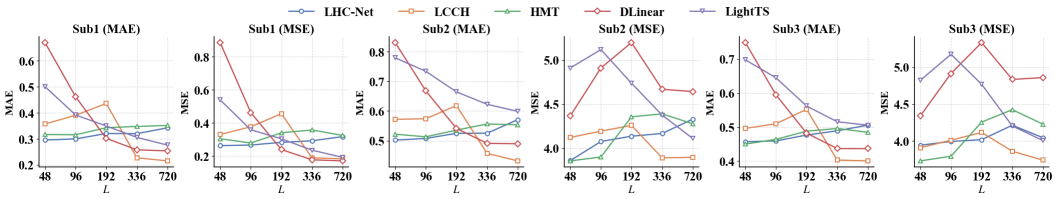

Figure 11: Sensitivity analysis of the lookback window length L across sub-datasets.

L.2 Impact of Lookback Window Length

To evaluate the dependency on historical observation length, we selected the top-5 performing baselines from the Sec.4.3 experiment. Fixing the prediction horizon at $H = 96$, we compared MSE/MAE fluctuations across different lookback window lengths $L \in \{48, 96, 192, 336, 720\}$ in Fig. 11.

Unlike models like DLinear that rely on long sequences to fit trends, LHC-Net maintains SOTA performance even with extremely short observation windows ($L = 48$). This is attributed to the semantic priors injected by LSI, which act as knowledge compensation to bridge the information scarcity of physical data in short windows. As L extends to 720, many baselines exhibit performance fluctuations or overfitting due to the introduction of excessive historical noise, whereas LHC-Net maintains a stable low-error trajectory. This confirms that the semantic retrieval mechanism of CRD effectively filters redundancy in long sequences, extracting only key features relevant to the current prediction. Consequently, LHC-Net is insensitive to variations in L, demonstrating robust adaptability for diverse application scenarios.

M Computational Efficiency and Industrial Scalability

To thoroughly evaluate the practical deployment viability of our proposed framework, we compare the full training/validation and inference time, parameter scale, and peak memory consumption of LHC-Net against four competitive baselines—Time-LLM, TimeMKG, HMT, and Crossformer—across various prediction horizons ($H \in \{96, 192, 336, 720\}$) on the private Sub1, Sub2, and Sub3 sub-datasets.

As summarized in Fig. 12, LHC-Net demonstrates a clear efficiency-effectiveness trade-off advantage. In terms of model complexity, LHC-Net contains only 15.9M parameters, which is markedly more lightweight than TimeMKG (38.0M), HMT (20.0M), Crossformer (94.8M), and especially Time-LLM (599.8M)—representing roughly a **37**$\times$ reduction in parameter count relative to the LLM-based Time-LLM baseline. Despite this compact architecture, LHC-Net consistently achieves the lowest or near-lowest total time (training/validation plus inference) among all compared methods across the three sub-datasets, typically completing a full cycle within 9–22 hours, whereas Time-LLM requires 60–78 hours due to its heavy LLM backbone. Given that industrial fermentation processes typically span hundreds of hours, this efficient turnaround time makes LHC-Net entirely practical for real-world pharmaceutical manufacturing applications, while also outperforming lighter baselines such as HMT and Crossformer, whose runtimes grow noticeably at longer horizons (e.g., $H = 336$ and $H = 720$).

Fig. 13 further reports the peak GPU memory footprint. Time-LLM incurs a substantial and rapidly escalating memory cost (from approximately 50 GiB at $H = 96$ to over 75 GiB at $H = 720$), reflecting the prohibitive resource demands of full LLM inference at scale. In contrast, LHC-Net maintains a low and stable memory profile comparable to TimeMKG at short horizons, and although its memory usage grows moderately with longer prediction horizons (reaching around 22–23 GiB at $H = 720$), it remains substantially below Crossformer (up to 29 GiB) and on par with or better than HMT, without ever approaching the memory ceiling imposed by Time-LLM. This favorable memory scalability, combined with its small parameter footprint, confirms that LHC-Net can be deployed on modest industrial-grade GPU hardware without requiring the expensive infrastructure typically associated with LLM-centric forecasting pipelines.

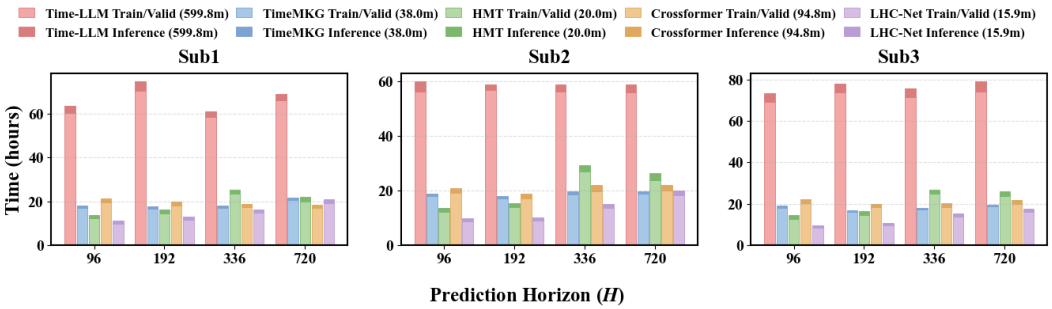

Figure 12: Total time allocation (hours) detailing training/validation versus inference durations for LHC-Net and baseline models (Time-LLM, TimeMKG, HMT, Crossformer) across the Sub1, Sub2, and Sub3 datasets, with corresponding parameter counts (M) annotated in the legend: Time-LLM (599.8M), TimeMKG (38.0M), HMT (20.0M), Crossformer (94.8M), and LHC-Net (15.9M).

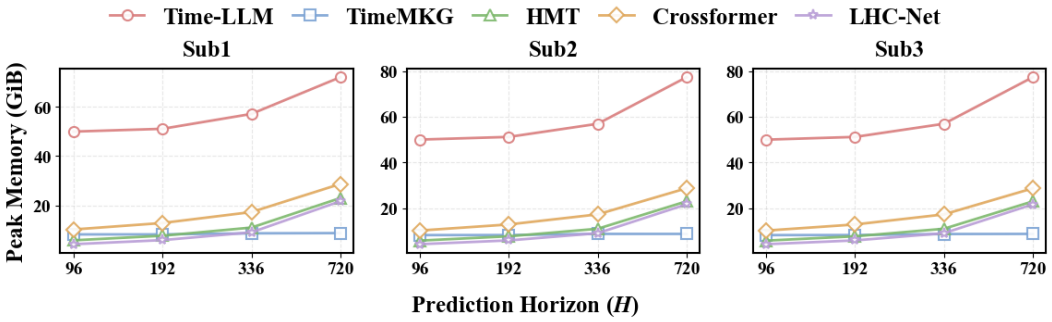

Figure 13: Peak memory consumption (GiB) of LHC-Net and baseline models across the Sub1, Sub2, and Sub3 datasets under varying prediction horizons.

Beyond raw computational speed and memory efficiency, LHC-Net exhibits strong industrial scalability and an economical deployment paradigm. The primary initial engineering investment lies merely in the construction of a domain-specific knowledge base, integrating daily factory operation logs with established scientific literature. Notably, our framework bypasses the computationally expensive fine-tuning typically required for LLMs—a factor that directly explains why Time-LLM's massive 599.8M-parameter backbone incurs disproportionately higher time and memory costs without a corresponding proportional gain in efficiency. Instead, LHC-Net seamlessly leverages prompt engineering alongside RAG to instruct an off-the-shelf LLM in autonomously constructing the knowledge-driven hypergraph, keeping the downstream forecasting backbone lightweight (15.9M parameters) while still injecting rich domain knowledge into the model.

This design inherently facilitates broad replicability. For other analogous biochemical scenarios, industrial practitioners can rapidly adapt our paradigm by building a localized, custom knowledge base, utilizing an off-the-shelf LLM for topology construction, and subsequently feeding the structured knowledge into downstream data-driven models for forecasting. Consequently, LHC-Net not only ensures scalable, lightweight, and memory-efficient deployment but also provides complete transparency and interpretability in its structural design.

N Exploration of Automated Latent Group Discovery

As indicated by our previous analyses shown in the Table 8, current LLMs cannot yet fully autonomously substitute the nuanced reasoning capabilities of frontline experts in highly specialized biochemical domains. From the perspective of practical industrial deployment, we argue that a hybrid paradigm represents the optimal solution to balance engineering robustness and model interpretability. Specifically, this paradigm entails combining low-effort manual definitions of macro-parameters

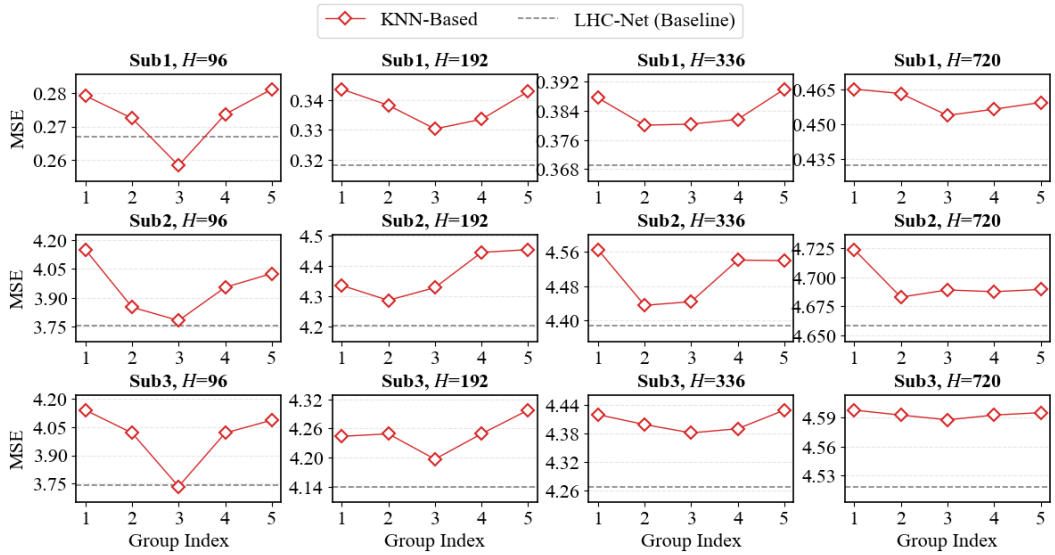

Figure 14: Predictive performance comparison between KNN-based automated grouping (varying the number of groups G) and the expert-guided LHC-Net across different prediction horizons (H).

(by experts) with high-complexity mining of micro-hypergraph associations (by LLMs), which are ultimately refined by data-driven dynamic corrections.

To further validate this design choice and explore the necessity of expert guidance, we conducted an experiment substituting the expert-defined latent mechanistic groups in LHC-Net with an automated discovery approach. Inspired by automated grouping strategies in existing literature Wang et al. (2024), we applied K-Nearest Neighbors (KNN) to automatically partition the sensor variables into functional clusters. We performed a sensitivity analysis on the number of automated latent groups, varying $G \in \{1, 2, 3, 4, 5\}$, and compared the predictive performance against our default expert-guided LHC-Net.

The comparative results across different prediction horizons are presented in Fig. 14. Interestingly, for short-term forecasting ($H = 96$), the automated KNN clustering method (specifically at $G = 3$) slightly outperforms the expert-defined parameters of LHC-Net. However, as the prediction horizon extends into long-term scenarios ($H = 192, 336, 720$), the expert-guided functional groups demonstrate significantly superior stability and lower prediction errors.

This performance divergence highlights a critical insight: while data-driven automated clustering can effectively capture local, short-term data correlations, it struggles to maintain global structural coherence over extended periods. In contrast, injecting deterministic domain knowledge to define macro-groups effectively constrains the hypothesis space, preventing the accumulation of non-stationary errors over time. This strongly supports our core proposition that the hybrid paradigm—leveraging expert knowledge for macro-structures and LLMs for micro-associations—is highly effective and essential for long-term, complex biochemical forecasting.

O Knowledge Base Construction and Prompt Design

LHC-Net leverages a robust industrial RAG pipeline to inject domain-specific priors for the initialization of the incidence matrix $\mathbf{H}$. To ensure a balance between precise terminology matching and semantic generalization, relevant knowledge concerning the target fermentation variables is extracted from public literature and internal factory Standard Operating Procedures (SOPs). This is achieved through a concurrent dual-pathway retrieval strategy—integrating BM25-based sparse retrieval with vector-based dense retrieval—followed by a strict ID-based deduplication process to maintain information fidelity. Furthermore, to mitigate risks such as context overflow and prompt

pollution, retrieved documents are isolated using structured XML-style tags (`[Knowledge]`) before being concatenated with the target variables (`[Input data]`) to form the final input to the LLM.

For further implementation details, including retrieval hyperparameters and the industrial knowledge base schema, please refer to `https://anonymous.4open.science/r/LHC-Net-8B41`.

P LIMITATIONS

While LHC-Net achieves promising results in modeling complex biochemical dynamics, it possesses certain limitations. First, the framework is specifically tailored for and evaluated on biopharmaceutical fermentation processes; its effectiveness and generalizability across broader multivariate time series domains (e.g., traffic, finance, weather) remain to be experimentally verified. Second, the LLM-guided hypergraph construction inherently relies on explicit semantic meanings and physical metadata. Consequently, LHC-Net cannot currently process datasets with anonymized, encrypted, or completely meaningless variable identifiers where semantic context is absent. Finally, although we explored fully automated, data-driven clustering for latent group discovery (detailed in **Appendix N**), achieving optimal predictive stability in long-term forecasting still necessitates a minor degree of frontline expert intervention to define macro-structures. Completely eliminating this reliance on human expertise without compromising structural coherence remains an important direction for future research.

Table 5: Comparison of baseline models on all sub-datasets of EPD. The **best** and second-best results for each row are highlighted. The Wins row counts the number of first-place finishes for each model on that dataset.

Model	H=96		H=192		H=336		H=720		Avg		Wins	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Sub1												
Crossformer	0.426	0.399	0.475	0.429	0.495	0.450	0.580	0.512	0.494	0.448	0	0
HMT	0.279	0.316	0.340	0.351	0.395	0.399	0.437	**0.440**	0.363	0.377	0	1
PatchTST	0.447	0.411	0.483	0.433	0.502	0.454	0.577	0.513	0.502	0.453	0	0
FEDformer	0.404	0.415	0.461	0.448	0.485	0.466	0.599	0.541	0.487	0.468	0	0
DLinear	0.463	0.462	0.468	0.467	0.475	0.477	0.531	0.520	0.484	0.482	0	0
CEGlo	0.976	0.712	0.924	0.697	0.890	0.699	0.969	0.749	0.940	0.714	0	0
LightTS	0.359	0.392	0.385	0.409	0.409	0.427	0.488	0.487	0.410	0.429	0	0
LCCH	0.377	0.391	0.409	0.414	0.466	0.465	0.609	0.549	0.465	0.455	0	0
Time-LLM	0.371	0.364	0.424	0.397	0.457	0.426	0.554	0.501	0.452	0.422	0	0
iTransformer	0.411	0.387	0.461	0.418	0.485	0.443	0.609	0.526	0.492	0.443	0	0
DHSL	0.292	0.319	0.338	0.356	0.373	0.387	0.457	0.456	0.365	0.380	0	0
TimeMKG	0.380	0.362	0.427	0.396	0.459	0.424	0.546	0.491	0.453	0.418	0	0
LHC-Net	**0.267**	**0.300**	**0.318**	**0.345**	**0.369**	**0.384**	**0.432**	0.442	**0.347**	**0.368**	**5**	**4**
Sub2												
Crossformer	4.592	0.568	5.912	0.658	6.680	0.705	7.438	0.781	6.156	0.678	0	0
HMT	3.901	0.513	4.494	0.567	4.584	0.597	4.689	**0.649**	4.417	0.582	0	1
PatchTST	4.566	0.658	6.111	0.610	6.760	0.642	7.576	0.705	6.253	0.654	0	0
FEDformer	5.242	0.642	6.463	0.710	6.879	0.657	7.934	0.856	6.630	0.716	0	0
DLinear	4.909	0.668	5.229	0.685	5.287	0.700	5.361	0.739	5.197	0.698	0	0
CEGlo	5.812	0.942	6.761	0.856	7.175	0.828	7.924	0.864	6.918	0.873	0	0
LightTS	5.117	0.734	5.437	0.748	5.478	0.765	5.495	0.795	5.381	0.761	0	0
LCCH	4.194	0.574	4.450	0.610	4.598	0.647	4.824	0.736	4.517	0.641	0	0
Time-LLM	4.318	0.564	6.048	0.647	6.694	0.697	7.278	0.763	6.085	0.668	0	0
iTransformer	3.860	0.559	5.741	0.646	6.638	0.699	7.360	0.776	5.900	0.670	0	0
DHSL	3.959	0.522	4.237	0.553	4.438	0.583	4.692	0.649	4.332	0.577	0	0
TimeMKG	3.931	0.542	5.863	0.630	6.512	0.679	7.225	0.761	5.883	0.653	0	0
LHC-Net	**3.753**	**0.507**	**4.201**	**0.545**	**4.386**	**0.583**	**4.658**	0.649	**4.249**	**0.571**	**5**	**4**
Sub3												
Crossformer	4.592	0.543	6.178	0.604	6.934	0.638	7.569	0.693	6.318	0.620	0	0
HMT	3.796	0.464	4.459	0.516	4.421	0.537	4.548	**0.561**	4.306	0.520	0	1
PatchTST	4.556	0.551	6.111	0.610	6.760	0.642	7.576	0.705	6.251	0.627	0	0
FEDformer	5.116	0.565	6.306	0.620	6.879	0.657	7.780	0.766	6.520	0.652	0	0
DLinear	4.912	0.595	5.131	0.612	5.137	0.623	5.188	0.656	5.092	0.622	0	0
CEGlo	5.539	0.834	6.761	0.856	7.175	0.828	7.924	0.864	6.850	0.846	0	0
LightTS	5.171	0.645	5.334	0.659	5.309	0.671	5.302	0.699	5.279	0.669	0	0
LCCH	4.012	0.510	4.313	0.544	4.455	0.574	4.670	0.641	4.363	0.567	0	0
Time-LLM	4.166	0.499	5.857	0.576	6.364	0.608	7.510	0.697	5.974	0.595	0	0
iTransformer	4.482	0.527	6.172	0.593	6.816	0.631	7.711	0.724	6.295	0.619	0	0
DHSL	3.959	0.463	4.188	0.492	4.341	0.515	4.578	0.580	4.266	0.513	0	0
TimeMKG	4.275	0.501	6.141	0.572	6.749	0.610	7.426	0.674	6.148	0.589	0	0
LHC-Net	**3.743**	**0.460**	**4.140**	**0.494**	**4.268**	**0.526**	**4.518**	0.574	**4.167**	**0.513**	**5**	**4**

Table 6: Comparison of baseline models on all sub-datasets of IndPenSim Dataset. The **best** and second-best results for each row are highlighted. The Wins row counts the number of first-place finishes for each model on that dataset.

Model	H=96		H=192		H=336		H=720		Avg		Wins	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Exhaust Gas												
Crossformer	**0.733**	0.347	**0.895**	0.422	1.002	0.477	1.176	0.557	0.952	0.451	2	0
HMT	0.812	0.326	0.958	0.417	0.993	**0.443**	1.148	0.541	0.978	0.432	0	1
PatchTST	1.022	0.421	1.340	0.561	1.650	0.698	2.113	0.877	1.531	0.639	0	0
FEDformer	1.219	0.483	1.695	0.600	2.272	0.728	2.153	0.853	1.835	0.666	0	0
DLinear	0.990	0.445	1.218	0.550	1.378	0.629	1.456	0.674	1.261	0.575	0	0
CEGlo	1.327	0.622	2.202	0.963	2.025	0.817	2.519	1.000	2.018	0.851	1	0
LightTS	1.017	0.419	1.077	0.494	1.157	0.548	1.312	0.618	1.141	0.520	0	0
LCCH	0.948	0.427	1.059	0.523	1.150	0.567	1.271	0.615	1.107	0.533	0	0
Time-LLM	1.251	0.376	3.005	0.499	2.834	0.613	2.290	0.663	2.345	0.538	0	0
iTransformer	1.358	0.407	2.379	0.542	1.792	0.663	2.175	0.844	1.926	0.614	0	0
DHSL	0.805	0.411	0.937	0.433	1.033	0.488	1.198	0.564	0.993	0.474	0	0
TimeMKG	1.118	0.424	1.537	0.555	1.839	0.695	2.214	0.864	1.677	0.635	0	0
LHC-Net	0.794	**0.314**	0.899	**0.374**	**0.982**	0.448	1.130	**0.527**	**0.951**	**0.416**	3	4
Vessel Control												
Crossformer	0.581	0.361	0.690	0.445	0.843	0.539	**0.883**	0.564	0.749	0.477	1	0
HMT	0.661	0.411	0.758	0.480	0.925	0.571	0.989	0.596	0.833	0.515	0	0
PatchTST	0.618	0.355	0.886	0.446	1.568	0.536	2.113	0.686	1.296	0.506	0	0
FEDformer	0.683	0.452	0.933	0.566	1.212	0.711	1.537	0.826	0.914	0.639	0	0
DLinear	0.624	0.439	0.815	0.540	0.957	0.626	1.071	0.695	0.867	0.575	0	0
CEGlo	0.766	0.501	1.451	0.786	1.364	0.730	1.742	0.885	1.331	0.726	0	0
LightTS	0.589	0.405	0.733	0.484	0.833	0.545	0.977	0.640	0.783	0.518	0	0
LCCH	0.642	0.445	0.798	0.533	0.902	0.599	1.031	0.661	0.843	0.560	0	0
Time-LLM	0.592	0.355	0.727	0.431	0.934	0.527	1.556	0.698	0.952	0.503	0	0
iTransformer	0.567	0.363	0.716	0.453	0.933	0.550	1.315	0.714	0.883	0.520	2	0
DHSL	0.535	0.381	0.695	0.459	0.779	0.514	0.930	0.607	0.734	0.490	0	0
TimeMKG	0.591	0.395	0.795	0.507	1.052	0.625	1.475	0.794	0.978	0.580	0	0
LHC-Net	**0.533**	**0.349**	**0.662**	**0.426**	**0.743**	**0.476**	0.884	**0.562**	**0.706**	**0.453**	4	5

Table 7: Comparison of LHC-Net against baseline model. An asterisk (*) indicates that LHC-Net significantly outperforms the baseline model ($p < 0.05$).

Dataset	*H*	LHC-Net (Ours)		HMT	
		MSE	**MAE**	**MSE**	**MAE**
Sub1	96	0.261-0.269*	0.299-0.301*	0.279-0.279	0.310-0.318
	192	0.314-0.319*	0.342-0.346*	0.340-0.342	0.335-0.356
	336	0.350-0.375*	0.375-0.387*	0.395-0.395	0.399-0.399
	720	0.424-0.456	0.439-0.451*	0.437-0.437	0.440-0.440
	avg	0.346-0.348*	0.367-0.369*	0.363-0.363	0.371-0.379
Sub2	96	3.650-4.071	0.507-0.507*	3.858-4.034	0.508-0.529
	192	4.094-4.532	0.544-0.552*	4.494-4.494	0.567-0.567
	336	4.359-4.395*	0.582-0.589*	4.584-4.584	0.597-0.597
	720	4.551-4.693	0.647-0.656	4.689-4.689	0.649-0.649
	avg	4.208-4.379*	0.570-0.575*	4.406-4.451	0.581-0.585
Sub3	96	3.702-3.871	0.458-0.462*	3.796-3.796	0.464-0.464
	192	4.075-4.343*	0.490-0.509*	4.459-4.459	0.516-0.516
	336	4.262-4.289*	0.526-0.527*	4.421-4.421	0.537-0.537
	720	4.503-4.569	0.572-0.581	4.548-4.548	0.561-0.561
	avg	4.136-4.267*	0.512-0.520*	4.306-4.306	0.519-0.521

Table 8: Robustness and feasibility analysis of LSI across 9 representative LLMs.

Dataset	Model	MAE ↓	QWK ↑	NDCG ↑	Time(s) ↓
Sub1	Gemini-3-pro	0.000	1.000	1.000	16.4
	Gemini-2.5-pro	0.213	0.915	0.954	18.9
	Gemini-2-flash	0.240	0.874	0.982	4.1
	GPT-5.2	0.120	0.955	0.999	5.2
	GPT-5	0.173	0.948	0.994	5.2
	GPT-4o	0.227	0.868	0.982	4.1
	DeepSeek-R1	0.147	0.943	0.984	27.8
	DeepSeek-V3.2	0.200	0.930	0.998	24.7
	DeepSeek-V3	0.080	0.973	0.998	13.5
Sub2	Gemini-3-pro	0.107	0.944	0.982	15.4
	Gemini-2.5-pro	0.173	0.798	0.982	17.4
	Gemini-2-flash	0.173	0.836	0.991	3.3
	GPT-5.2	0.067	0.978	0.992	5.6
	GPT-5	0.213	0.867	0.985	4.9
	GPT-4o	0.267	0.780	0.965	3.5
	DeepSeek-R1	0.187	0.873	0.987	45.3
	DeepSeek-V3.2	0.213	0.884	0.969	26.4
	DeepSeek-V3	0.240	0.854	0.954	13.3
Sub3	Gemini-3-pro	0.133	0.947	0.984	14.9
	Gemini-2.5-pro	0.240	0.886	0.956	18.5
	Gemini-2-flash	0.227	0.838	0.985	3.0
	GPT-5.2	0.093	0.960	0.987	5.5
	GPT-5	0.147	0.864	0.992	4.9
	GPT-4o	0.227	0.849	0.980	3.9
	DeepSeek-R1	0.173	0.915	0.979	43.3
	DeepSeek-V3.2	0.280	0.861	0.957	23.1
	DeepSeek-V3	0.413	0.712	0.960	15.5