TD-Hyper: Time-Density Guided Hypergraph Learning for Clinical Irregular Multivariate Time Series

1st Guohui Ding *School of Computer Science*
Shenyang Aerospace University
Shenyang, China
dingguohui@sau.edu.cn

2nd Zejun Zhang *School of Computer Science*
Shenyang Aerospace University
Shenyang, China
zhangzejun@stu.sau.edu.cn

3rd Jinwei Wang *Department of Medicine*
Peking University First Hospital
Beijing, China
wangjinwei@bjmu.edu.cn

***Abstract*—Irregular multivariate time series (IMTS) are ubiquitous in clinical settings, where observations are asynchronously recorded at non-uniform intervals across variables. Most existing methods mitigate temporal irregularity through resampling, interpolation, or alignment, implicitly treating sampling density as a nuisance rather than an informative signal. However, in clinical practice, measurement frequency itself reflects patient state, giving rise to alternating dense and sparse sampling regimes along the timeline. In this paper, we propose TD-Hyper, a density-conditioned hypergraph framework that treats sampling density as a first-class structural variable governing both temporal neighborhood construction and information propagation. TD-Hyper models asynchronous observations with soft adaptive temporal hyperedges and introduces a density-gated dual-mode message passing mechanism to differentiate dense and sparse sampling regimes without resampling or explicit temporal alignment. Experiments on multiple clinical IMTS benchmarks demonstrate that TD-Hyper consistently outperforms strong baselines while maintaining a moderate computational footprint.**

***Index Terms*—Irregular Multivariate Time Series, Time-Density, Density-Aware Hypergraph, Soft Temporal Hyperedges, Clinical Time Series**

I. Introduction

With the rapid digitalization of healthcare systems, large-scale clinical time series play an important role in tasks such as early warning, risk stratification, and outcome prediction. In practice, these data are typically collected as *irregular multivariate time series* (IMTS), where observations occur at non-uniform time intervals and different variables are measured asynchronously. This dual irregularity challenges conventional temporal models that assume regularly sampled and temporally aligned inputs. Existing approaches address irregularity through resampling, interpolation, discretization, or continuous-time attention mechanisms. Although these methods mitigate temporal misalignment, they generally treat irregularity—especially sampling density—as a nuisance to be compensated for, aiming to reconstruct a latent regular timeline or suppress artifacts caused by non-uniform sampling.

However, in clinical practice, sampling density is rarely random. Measurement frequency is often driven by patient condition: unstable states trigger dense monitoring, while stable periods are sparsely observed. Consequently, sampling density encodes meaningful signals about physiological dynamics and clinical decision processes, and induces heterogeneous *information regimes*. Dense regions provide abundant but potentially redundant temporal evidence, where excessive aggregation may lead to oversmoothing, whereas sparse regions suffer from weak temporal support and require stronger cross-variable corroboration. Treating these regimes uniformly may therefore result in suboptimal temporal representations.

Motivated by this observation, we propose **TD-Hyper**, a **time-density guided hypergraph learning framework** for clinical IMTS. Instead of enforcing hard temporal alignment, TD-Hyper represents each observation as a node and constructs a density-aware hypergraph with two complementary structures: (i) *variable hyperedges* capturing within-variable temporal continuity, and (ii) *soft temporal hyperedges* that form overlapping temporal neighborhoods across asynchronous observations, enabling continuous temporal aggregation without resampling. TD-Hyper further treats sampling density as a first-class structural signal: local density parameterizes both the memberships and the importance of temporal hyperedges, allowing aggregation scopes to adapt to heterogeneous sampling regimes. To account for the asymmetric information conditions between dense and sparse regions, we additionally introduce a *density-gated dual-mode message passing mechanism*, which balances time-centric aggregation in dense regions with enhanced cross-variable reasoning in sparse regions.

Our main contributions are summarized as follows:

- We reconceptualize sampling density in IMTS as a structural signal that defines heterogeneous information regimes rather than a nuisance factor.
- We propose soft adaptive temporal hyperedges for continuous and uncertainty-aware temporal aggregation over irregular and asynchronous observations.
- We introduce density-aware temporal hyperedge parameterization that embeds local sampling density into the hypergraph structure.
- We design a density-gated dual-mode message passing strategy to balance temporal and cross-variable information propagation under heterogeneous sampling regimes.

II. Related Work

A. Irregular Multivariate Time Series Modeling

Existing research on irregular multivariate time series (IMTS) focuses on tasks such as classification [1]–[3], interpolation [1], and forecasting [4], [5]. From a modeling perspective, existing approaches can be broadly categorized into resampling-based methods and raw-time methods.

Resampling-based approaches normalize irregular observations onto a uniform grid through interpolation or kernel-based aggregation [6], [7]. While effective, these operations may smooth or distort the original sampling patterns and weaken the density information inherent in raw sequences.

Raw-time methods instead operate directly on irregular observations. Representative approaches include RNN variants designed for uneven time gaps [8], attention models with continuous-time embeddings [1], [3], and neural ODE-based frameworks for modeling continuous dynamics [9]. To further capture cross-variable dependencies, recent studies incorporate attention mechanisms [10] or graph-based structures [4], [5], [11]. Despite strong empirical performance, most existing IMTS models treat sampling density mainly as an artifact reflected by time gaps or missing values, rather than an explicit structural signal governing temporal aggregation and information propagation.

III. Problem Definition

Let $\mathcal{D} = \{(\mathcal{S}_i, y_i)\}_{i=1}^{N}$ denote an irregular multivariate time series (IMTS) dataset for classification, where $y_i \in \mathcal{Y} = \{1, \dots, C\}$. Each sample is a set of timestamped observations $\mathcal{S}_i = \{(t_{ij}, z_{ij}, u_{ij})\}_{j=1}^{M_i}$, where $t_{ij} \in \mathbb{R}_+$ is the timestamp, $z_{ij} \in \mathbb{R}$ the observed value, and $u_{ij} \in \{1, \dots, U\}$ the variable index. IMTS are irregular (non-uniform time gaps) and asynchronous (variables have distinct timestamp sets). The goal is to learn a classifier F_θ that predicts $\hat{y}_i = \arg\max_{c \in \mathcal{Y}} \big(\mathrm{softmax}(F_\theta(\mathcal{S}_i))\big)_c$.

IV. Methodology

We now present *TD-Hyper*, a time-density guided hypergraph framework for clinical IMTS that elevates sampling density to a structural signal governing both temporal neighborhood construction and information propagation. Sec. IV-A formalizes the density-aware hypergraph representation and its basic components. Sec. IV-B introduces soft temporal hyperedges that replace hard binary time windows with graded, overlapping temporal neighborhoods, enabling continuous temporal aggregation over irregular time axes, as illustrated in Fig. 1(a). Sec. IV-C then defines a normalized time-density signal $\rho(t)$ that conditions temporal neighborhood scale and strength, inducing adaptive kernel bandwidths under heterogeneous sampling patterns. Sec. IV-D details how this density signal further shapes hyperedge weighting and updating, embedding sampling density directly into the structural learning process. Finally, Sec. IV-E presents a density-gated message passing mechanism that conditions information routing on local sampling density, as illustrated in Fig. 1(b).

A. Density-Aware Hypergraph for IMTS

Instead of relying on padding or patch-based discretization, TD-Hyper reframes irregular observations as nodes in a density-aware hypergraph, preserving intrinsic sampling-density cues while enabling structured and efficient representation learning. As illustrated in Fig. 1, for each IMTS sample $S_i = \{(t_{ij}, z_{ij}, u_{ij})\}_{j=1}^{M_i}$, we construct a density-aware hypergraph $\mathcal{G}_i := \big(\mathcal{V}_i, \mathcal{E}_i, \mathbf{W}_i\big)$, where the observation set $\mathcal{V}_i := \{v_j \mid j = 1, \dots, M_i\}$ forms the node space and is connected by two complementary families of hyperedges: temporal hyperedges $\mathcal{E}_i^{\mathrm{time}}$, which encode local temporal structure, and variable hyperedges $\mathcal{E}_i^{\mathrm{var}}$, which capture within-variable dependencies.

a) Variable hyperedges: TD-Hyper groups observations associated with the same variable into a variable hyperedge e_u, which serves as a structural carrier of within-variable dependencies across irregular time. Formally, variable hyperedges are encoded through a binary incidence relation:

$$(\mathbf{H}_i^U)_{j,u} = \begin{cases} 1, & v_j \in e_u, \\ 0, & v_j \notin e_u, \end{cases} \qquad u = 1, \dots, U, \tag{1}$$

where $\mathbf{H}_i^U \in \{0,1\}^{M_i \times U}$ denotes the variable–incidence matrix.

b) Soft temporal hyperedges: Each observed timestamp τ_t anchors a temporal hyperedge e_t that induces a local temporal context by connecting nearby observations with graded memberships. Formally, the soft temporal incidence is defined as

$$(\widetilde{\mathbf{H}}_i^T)_{j,t} \in [0,1], \tag{2}$$

where multiple nonzero entries in each row allow an observation v_j to participate in several overlapping temporal hyperedges with different strengths. Unlike hard binary incidence relations, this construction yields continuous and uncertainty-aware temporal neighborhoods, which form the structural substrate for density-conditioned temporal aggregation. The detailed construction is described in Sec. IV-B.

c) Embedding initialization.: Observation nodes are initialized from raw values $\mathbf{Z}_i$ through a shallow nonlinear projection, yielding initial node representations

$$\mathbf{V}_i = \mathrm{ReLU}\big(\mathrm{FF}_{\mathrm{obs}}(\mathbf{Z}_i)\big) \in \mathbb{R}^{M_i \times P_{\mathrm{obs}}}. \tag{3}$$

Temporal hyperedges are endowed with both positional and density-aware embeddings. Specifically, each temporal hyperedge is encoded using a sinusoidal time representation (after a linear projection) to capture its temporal location, concatenated with a learned density embedding that encodes local sampling regimes:

$$\mathbf{E}_i^{\mathrm{time}} = \big[\, \sin\big(\mathrm{FF}_{\mathrm{time}}(\boldsymbol{\tau}_i)\big) \,\big\|\, \mathrm{FF}_\rho(\boldsymbol{\rho}_i) \,\big] \in \mathbb{R}^{T_i \times (P_{\mathrm{time}} + P_\rho)}. \tag{4}$$

Variable hyperedges are initialized as learnable embeddings, allowing the model to capture latent variable-specific factors:

$$\mathbf{E}_i^{\mathrm{var}} = \mathrm{ReLU}(\mathbf{W}_{\mathrm{var}}) \in \mathbb{R}^{U \times P_{\mathrm{var}}}. \tag{5}$$

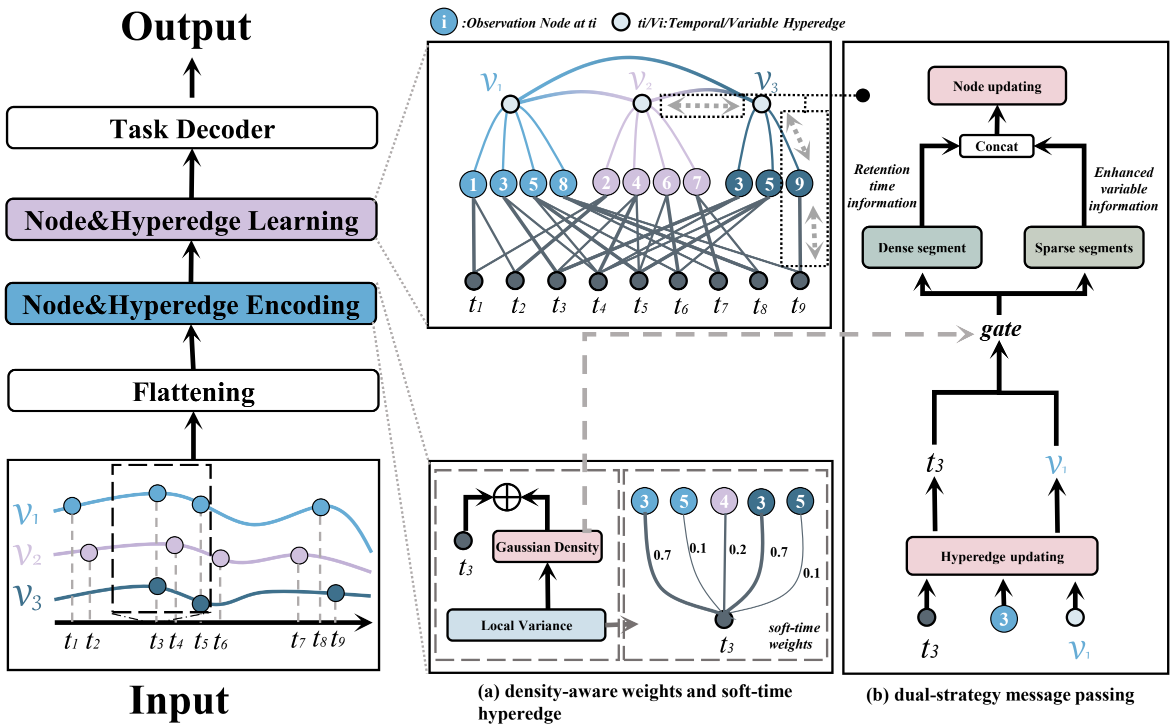

Fig. 1. The architecture of TD-Hyper. It first converts the input IMTS sample into a valid hypergraph representation. (a) The Gaussian kernel density is calculated by computing the local variance of each observation point, which is used to update the hyperedge weights and generate soft temporal hyperedges. (b) Hyperedge embeddings and node embeddings are updated through Dual-mode message passing.

B. Soft Temporal Hyperedges

To model temporal neighborhoods under irregular sampling, TD-Hyper replaces binary temporal incidence with soft temporal hyperedges that provide continuous, distance-aware memberships.

Given S_i, let observed timestamps serve as anchors $T_i = \{\tau_1, \ldots, \tau_{T_i}\}$. For each $\tau \in T_i$, we define a temporal hyperedge with Gaussian-decayed memberships:

$$\begin{aligned} \widetilde{H}_i^T(j,\tau) \;&=\; \exp\Big(-\tfrac{1}{2}\big(\tfrac{t_{ij}-\tau}{\sigma_i(\tau)}\big)^2\Big)\cdot \mathbb{I}\big(|t_{ij}-\tau| \le \Delta\big), \\ &\sum_{j=1}^{M_i} \widetilde{H}_i^T(j,\tau) = 1, \end{aligned} \tag{6}$$

This yields a soft temporal incidence matrix $\widetilde{\mathbf{H}}_i^T \in [0,1]^{M_i \times T_i}$ with column-wise normalization.

Each node can participate in multiple hyperedges with different strengths, inducing overlapping temporal neighborhoods. A temporal hyperedge anchored at τ performs local aggregation:

$$\mathbf{h}_\tau = \sum_{j=1}^{M_i} \widetilde{H}_i^T(j,\tau)\mathbf{X}_j. \tag{7}$$

The bandwidth $\sigma_i(\tau)$ controls neighborhood locality, while normalization ensures stable and comparable aggregation across anchors.

C. Density-Aware Parameterization of Temporal Hyperedges

While soft temporal hyperedges define continuous temporal neighborhoods, they do not account for heterogeneous sampling density. TD-Hyper addresses this by introducing a density signal to parameterize both temporal memberships and hyperedge strengths.

For each anchor τ, we estimate local density by combining sampling rate and value variability within a window $[\tau - \Delta_\rho, \tau + \Delta_\rho]$:

$$\begin{aligned} r_i(\tau) &= \frac{1}{2\Delta_\rho}\sum_{j=1}^{M_i} \mathbb{I}(|t_{ij}-\tau| \le \Delta_\rho) \\ s_i(\tau) &= \mathrm{Var}(\{\tilde{z}_{ij} : |t_{ij}-\tau| \le \Delta_\rho\}) \end{aligned} \tag{8}$$

$$\rho_i(\tau) = (1-\alpha_\rho)\, r_i(\tau) + \alpha_\rho\, \beta\, s_i(\tau), \tag{9}$$

We normalize $\rho_i(\tau)$ and map it to a density-dependent bandwidth:

$$\begin{aligned} \bar{\rho}_i(\tau) &= \frac{\rho_i(\tau) - \min_{\tau'} \rho_i(\tau')}{\max_{\tau'} \rho_i(\tau') - \min_{\tau'} \rho_i(\tau') + \varepsilon}, \\ \sigma_i(\tau) &= \sigma_{\min} + (\sigma_{\max} - \sigma_{\min})\big(1 - \bar{\rho}_i(\tau)\big). \end{aligned} \tag{10}$$

Thus, dense regions yield smaller bandwidths, while sparse regions induce wider temporal neighborhoods. In addition, temporal hyperedge weights are defined as

$$w_i^{(\mathrm{time})}(\tau) = \bar{\rho}_i(\tau), \quad \mathbf{W}_i^T = \mathrm{diag}\big(\{w_i^{(\mathrm{time})}(\tau)\}\big). \tag{11}$$

Hence, density jointly controls membership sharpness (via $\sigma_i(\tau)$) and aggregation strength (via $w_i^{(\mathrm{time})}(\tau)$).

D. Hyperedge Update

After initializing node, temporal-hyperedge, and variable-hyperedge embeddings, TD-Hyper updates hyperedges via attention-based message passing. For temporal hyperedges, queries are instantiated from the current temporal-hyperedge embeddings, while keys and values are formed by concatenating observation-node features with variable-hyperedge embeddings. A logarithmic bias derived from the column-normalized soft temporal incidence encodes the local temporal neighborhood structure and is added to the attention logits.

Formally, multi-head attention is applied to update temporal hyperedges, followed by a residual connection and a feed-forward projection:

$$\mathbf{E}_{\ell+1}^{(\mathrm{time})} = \mathbf{E}_{\ell}^{(\mathrm{time})} + \mathrm{ReLU}\big(\mathrm{FFO}(\mathbf{O}^{(\mathrm{time})})\big). \tag{12}$$

Variable hyperedges are updated analogously, with queries drawn from variable-hyperedge embeddings and keys and values formed by concatenating observation-node features with the updated temporal-hyperedge embeddings. The corresponding structural prior is provided by the variable–observation incidence matrix, which biases attention toward valid memberships.

E. Dual-Mode Message Passing

TD-Hyper employs a density-gated dual-mode message passing mechanism to adapt aggregation under heterogeneous sampling patterns.

For each node at time t_n, a density-dependent gate is defined as

$$g_n = \sigma(\alpha_\rho \rho_i(t_n)), \tag{13}$$

which balances contributions from temporal and variable hyperedges.

Temporal and variable messages are first aggregated from their respective hyperedges, and then fused with the node representation via gated integration:

$$X^{(\ell+1)} = \Big(X^{(\ell)} + \mathrm{ReLU}\big(W_{h2n}[O^{(\ell)} \| \tilde{T}^{(\ell)} \| \tilde{U}^{(\ell)}]\big)\Big) \odot m. \tag{14}$$

Here, g_n adaptively controls the relative influence of temporal versus variable information, enabling density-dependent routing of message passing.

V. Experiment

A. Experimental Setting

1) DATASETS: Five widely studied medical IMTS datasets are used in the experiments, namely P19 [12], PhysioNet [13], MIMIC-IV [14] , MIMIC-III [15] and P12 [16] where PhysioNet is a reduced version of P12 considered by prior work [1].

2) BASELINES: We compare our approach with state-of-the-art methods for irregular time series, including HyperIMTS [17], HI-Patch [18], GRU-D [8], mTAN [1], SeFT [3], GraFITi [4], Raindrop [19], tPatchGNN [5], Warpformer [2]. In addition, we compare our approach with state-of-the-art methods for regular time series, namely MICN [20], Ada_MSHyper [21], TiDE [22], TimeMixer [23]. The baselines are implemented using the official codebases whenever available. We follow the recommended hyperparameter settings in the original papers and further tune key hyperparameters on the validation sets to ensure fair comparison.

3) EVALUATION SETUP: For the classification task, we follow the method described in Harutyunyan et al. [24] and divide the dataset into three parts for training, validation, and testing with the ratio of 70%,15%,15% on the MIMIC-III and MIMIC-IV datasets. For the remaining three datasets, we adhered to Zhang et al.'s approaches [19], and the ratio of training, validation, and testing set is 8:1:1. We measure the classification performance with the Area Under the Receiver Operating Characteristic Curve (AUROC) and Area Under the Precision-Recall Curve (AUPRC) since all the five datasets are binary classification datasets with highly imbalanced class distribution.

B. Implementation Details

We adopt the Adam optimizer with an initial learning rate of 0.001. We apply early stopping and terminate training if the validation loss does not improve for 10 consecutive epochs. All experiments are conducted using five random seeds, and the results are reported as the mean and standard deviation across these runs. The models are implemented using the PyTorch library and trained on a single GeForce RTX-4090 24GB GPU.

C. Main Results

Table I summarizes the classification performance across the five benchmark datasets; the best scores are shown in bold and the second-best are underlined. TD-Hyper achieves the strongest overall results, attaining the top AUROC on every dataset and either the best or a very close second-best AUPRC. Compared to the HyperIMTS model, which constructs a hypergraph for the entire irregular sequence, TD-Hyper demonstrates significant performance improvements by learning the raw density information in the sequence through density-aware hyperedges. TD-Hyper also demonstrates significant improvement over TimeMixer, a strong regular-time-series model. In addition, we observe that some models for regular time series outperform some IMTS models, which also reflects the necessity of comprehensive evaluation of regular and irregular time series models.

D. Ablation Study

We evaluate **TD-Hyper** and three ablated variants on the P19 and P12 datasets: (i) **w/o STE** — soft temporal hyperedges are disabled and temporal memberships become hard/binary; (ii) **w/o DATH** — density-aware hyperedge coefficients are removed and replaced with uniform weights, thus ignoring raw sampling-density cues; and (iii) **w/o DMP** — the Dual-mode message passing is disabled and replaced by a single shared propagation operator. The results are summarized in Table II.

TABLE I
METHOD BENCHMARKING ON IMTS CLASSIFICATION. THE BEST RESULTS ARE HIGHLIGHTED IN **BOLD**, AND THE SECOND-BEST RESULTS ARE UNDERLINED. THE RESULTS ARE IN THE FORM OF (MEAN ± STD %).

Category	Methods	P19		PhysioNet		MIMIC-III		P12		MIMIC-IV	
		AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
IRREGULAR	HyperIMTS	92.2 ± 0.4	**66.2 ± 0.9**	83.3 ± 1.6	51.2 ± 1.8	82.7 ± 1.0	47.2 ± 1.8	85.3 ± 0.3	49.6 ± 1.4	82.9 ± 1.1	47.5 ± 1.7
	HI-Patch	92.2 ± 0.5	60.9 ± 2.4	83.3 ± 1.1	49.5 ± 1.6	82.9 ± 0.7	47.1 ± 1.0	86.2 ± 0.6	**51.6 ± 1.5**	83.1 ± 0.8	47.3 ± 1.1
	GRU-D	90.8 ± 0.8	63.3 ± 1.2	78.6 ± 2.4	41.4 ± 2.3	76.2 ± 0.4	37.1 ± 0.7	83.0 ± 1.0	41.5 ± 0.7	75.9 ± 0.5	36.8 ± 0.8
	mTAN	80.7 ± 1.1	26.0 ± 2.2	80.2 ± 1.6	44.4 ± 2.7	75.4 ± 1.7	37.6 ± 2.2	84.1 ± 1.3	47.7 ± 2.7	75.6 ± 1.6	37.0 ± 2.1
	SeFT	85.4 ± 1.1	59.9 ± 1.1	63.9 ± 2.2	22.7 ± 2.9	68.5 ± 1.1	24.9 ± 0.6	67.4 ± 1.5	20.2 ± 1.2	68.0 ± 1.2	25.1 ± 0.7
	GraFITi	91.7 ± 1.0	63.6 ± 1.0	80.3 ± 3.6	47.5 ± 2.2	83.8 ± 0.3	49.0 ± 0.7	85.6 ± 0.4	50.1 ± 0.9	83.6 ± 0.4	49.2 ± 0.6
	Raindrop	89.1 ± 0.5	62.3 ± 0.4	78.0 ± 2.2	38.4 ± 3.6	76.6 ± 0.5	38.1 ± 0.8	82.2 ± 0.5	42.7 ± 1.1	76.3 ± 0.6	38.4 ± 0.9
	tPatchGNN	88.4 ± 0.5	59.7 ± 1.4	75.0 ± 4.3	35.7 ± 6.7	– – –	– – –	80.7 ± 1.9	42.1 ± 3.4	75.4 ± 0.5	37.1 ± 1.0
	Warpformer	92.2 ± 0.6	60.1 ± 1.4	84.2 ± 1.6	49.5 ± 4.1	83.3 ± 1.0	48.4 ± 2.3	85.9 ± 0.5	51.0 ± 1.5	83.1 ± 1.1	48.6 ± 2.2
REGULAR	MICN	82.9 ± 0.5	46.0 ± 3.2	60.0 ± 1.9	20.2 ± 2.3	69.3 ± 0.6	31.5 ± 0.8	71.2 ± 1.9	29.6 ± 2.1	69.6 ± 0.7	31.2 ± 0.9
	Ada_MSHyper	81.1 ± 1.3	36.9 ± 2.1	67.2 ± 2.6	25.3 ± 3.4	70.5 ± 1.0	27.1 ± 1.5	74.3 ± 0.9	28.3 ± 1.0	70.2 ± 1.1	27.5 ± 1.4
	TiDE	85.6 ± 1.5	50.4 ± 4.9	76.9 ± 1.7	33.4 ± 1.5	76.7 ± 0.6	35.5 ± 0.7	79.4 ± 0.8	40.2 ± 1.6	76.9 ± 0.7	35.3 ± 0.8
	TimeMixer	87.5 ± 0.7	54.0 ± 0.6	76.7 ± 2.2	34.6 ± 4.7	77.1 ± 0.8	38.0 ± 1.8	82.6 ± 0.8	41.8 ± 1.4	76.8 ± 0.9	38.2 ± 1.7
	TD-Hyper	**93.3 ± 0.2**	65.1 ± 1.2	**84.8 ± 1.5**	**51.4 ± 2.6**	**84.6 ± 0.8**	**50.7 ± 1.5**	**86.6 ± 0.2**	49.7 ± 1.4	**84.7 ± 0.7**	**51.0 ± 1.4**

TABLE II
ABLATION RESULTS OF TD-HYPER ON TWO DATASETS. THE RESULTS ARE PRESENTED IN THE FORM OF (MEAN ± STD).

Methods	P19		P12	
	AUROC	AUPRC	AUROC	AUPRC
TD-Hyper	**93.3 ± 0.2**	**65.1 ± 1.2**	**86.6 ± 0.2**	**49.7 ± 1.4**
w/o STE	92.8 ± 0.8	64.9 ± 1.1	86.1 ± 0.2	49.1 ± 1.1
w/o DATH	92.8 ± 0.4	64.6 ± 0.3	86.2 ± 0.5	48.7 ± 2.5
w/o DMP	93.0 ± 0.4	65.0 ± 0.5	86.0 ± 0.6	48.4 ± 2.4

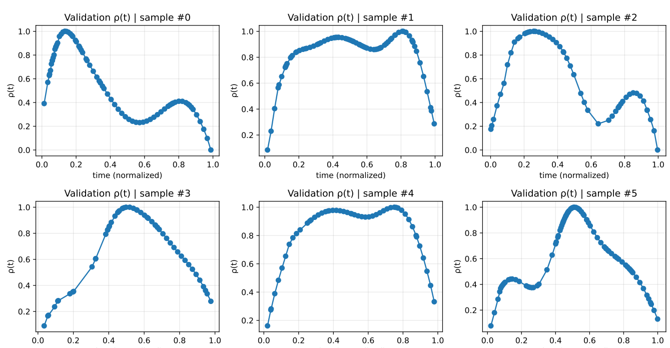

Fig. 2. Sampling-density profiles of representative validation samples from the P12 dataset. The curves show the normalized local sampling density $\rho_i(t)$ over time, highlighting substantial temporal variation in measurement frequency across patients and within individual trajectories.

E. Effect of Density

Fig. 2 visualizes the temporal evolution of sampling density for representative IMTS samples. We observe pronounced heterogeneity both across subjects and within individual trajectories, with dense bursts interleaved with sparsely sampled periods. This variability reflects realistic clinical monitoring patterns and supports our motivation to explicitly model sampling density as a structural signal rather than treating it as noise.

Fig. 3 examines the sensitivity of TD-Hyper to the local temporal window size Δ. We observe a clear interior optimum on both datasets: overly small windows under-aggregate cross-variable context, while overly large windows dilute temporally local patterns.

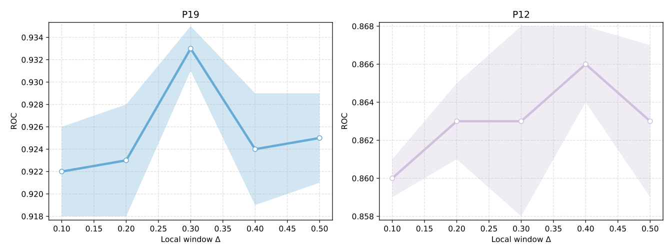

Fig. 3. Effect of the local temporal window size Δ on model performance. The figure reports AUROC as a function of Δ on the P12 and P19 datasets, illustrating the sensitivity of TD-Hyper to the temporal aggregation scope.

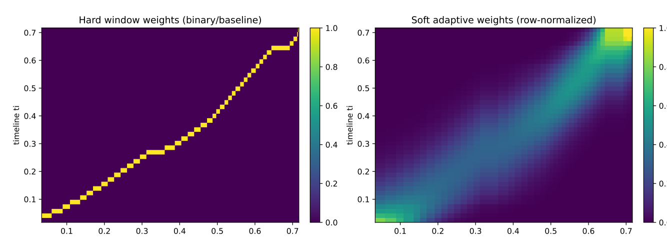

Fig. 4. Hard versus soft temporal hyperedges. Hard windows induce abrupt neighborhood changes: observations with negligible temporal differences may be either fully connected or completely disconnected depending on window boundaries. In contrast, soft temporal hyperedges assign graded memberships that decay smoothly with temporal distance, allowing overlapping temporal neighborhoods and avoiding discontinuous structural shifts.

As illustrated in Fig. 4, hard temporal windows enforce binary memberships based on exact or discretized timestamps, which can introduce abrupt boundary effects under irregular sampling. In contrast, the proposed soft adaptive temporal memberships assign continuous, distance-based weights within a local window, allowing temporally adjacent but asynchronous observations to contribute proportionally. This design preserves temporal uncertainty while facilitating cross-variable aggregation without resampling.

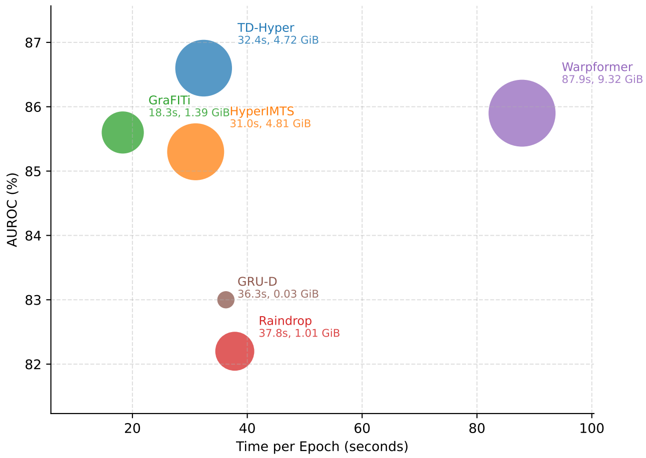

Fig. 5. Performance–efficiency trade-off on the P12 dataset. The x-axis denotes average training time per epoch, the y-axis denotes AUROC, and bubble size corresponds to peak GPU memory usage. All methods are evaluated with batch size = 32 under identical training settings.

F. Computational Complexity

Let M denote the number of observations, T the number of temporal hyperedges (timestamps), and d the embedding dimension. TD-Hyper performs attention-based message passing only within density-adaptive local temporal neighborhoods. Assuming a bounded average neighborhood size w, the per-layer complexity is $\mathcal{O}(T \cdot w \cdot d + M \cdot d)$.

which scales linearly with the number of irregular observations. Therefore, TD-Hyper avoids the $\mathcal{O}(M^2)$ global attention cost and remains scalable for long and irregular clinical time series, consistent with the empirical efficiency results in Fig. 5.

VI. Conclusion

This paper introduced TD-Hyper, a density-conditioned hypergraph framework for irregular multivariate time series that treats sampling density as a first-class structural signal rather than a byproduct of irregularity. Experiments on multiple clinical benchmarks demonstrate consistent performance improvements with moderate computational cost. Despite its advantages, TD-Hyper introduces additional modeling complexity and relies on a handcrafted density formulation, which may limit scalability in extremely long or ultra-dense sequences or in domains where sampling density is weakly correlated with information reliability. Future work will explore more scalable hypergraph constructions, learning-based density representations, and extensions of the density-conditioned paradigm to other irregular domains and tasks.

References

[1] S. N. Shukla and B. M. Marlin, "Multi-time attention networks for irregularly sampled time series," *arXiv preprint arXiv:2101.10318*, 2021.

[2] J. Zhang, S. Zheng, W. Cao, J. Bian, and J. Li, "Warpformer: A multi-scale modeling approach for irregular clinical time series," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 3273–3285, 2023.

[3] M. Horn, M. Moor, C. Bock, B. Rieck, and K. Borgwardt, "Set functions for time series," in *International Conference on Machine Learning*, pp. 4353–4363, PMLR, 2020.

[4] V. K. Yalavarthi, K. Madhusudhanan, R. Scholz, N. Ahmed, J. Burchert, S. Jawed, S. Born, and L. Schmidt-Thieme, "Grafiti: Graphs for forecasting irregularly sampled time series," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, pp. 16255–16263, 2024.

[5] W. Zhang, C. Yin, H. Liu, X. Zhou, and H. Xiong, "Irregular multivariate time series forecasting: A transformable patching graph neural networks approach," in *Forty-first International Conference on Machine Learning*, 2024.

[6] Y. Wu, J. Ni, W. Cheng, B. Zong, D. Song, Z. Chen, Y. Liu, X. Zhang, H. Chen, and S. B. Davidson, "Dynamic gaussian mixture based deep generative model for robust forecasting on sparse multivariate time series," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, pp. 651–659, 2021.

[7] Q. Tan, M. Ye, G. L.-H. Wong, and P. C. Yuen, "Cooperative joint attentive network for patient outcome prediction on irregular multi-rate multivariate health data.," in *IJCAI*, pp. 1586–1592, 2021.

[8] Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, "Recurrent neural networks for multivariate time series with missing values," *Scientific reports*, vol. 8, no. 1, p. 6085, 2018.

[9] M. Schirmer, M. Eltayeb, S. Lessmann, and M. Rudolph, "Modeling irregular time series with continuous recurrent units," in *International conference on machine learning*, pp. 19388–19405, PMLR, 2022.

[10] S. Y. Jhin, M. Jo, T. Kong, J. Jeon, and N. Park, "Ace-node: Attentive co-evolving neural ordinary differential equations," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pp. 736–745, 2021.

[11] Y. Luo, Z. Liu, L. Wang, B. Wu, J. Zheng, and Q. Ma, "Knowledge-empowered dynamic graph network for irregularly sampled medical time series," *Advances in Neural Information Processing Systems*, vol. 37, pp. 67172–67199, 2024.

[12] M. A. Reyna, C. S. Josef, R. Jeter, S. P. Shashikumar, M. B. Westover, S. Nemati, G. D. Clifford, and A. Sharma, "Early prediction of sepsis from clinical data: the physionet/computing in cardiology challenge 2019," *Critical care medicine*, vol. 48, no. 2, pp. 210–217, 2020.

[13] I. Silva, G. Moody, D. J. Scott, L. A. Celi, and R. G. Mark, "Predicting in-hospital mortality of icu patients: The physionet/computing in cardiology challenge 2012," in *2012 computing in cardiology*, pp. 245–248, IEEE, 2012.

[14] A. E. Johnson, L. Bulgarelli, L. Shen, A. Gayles, A. Shammout, S. Horng, T. J. Pollard, S. Hao, B. Moody, B. Gow, *et al.*, "Mimic-iv, a freely accessible electronic health record dataset," *Scientific data*, vol. 10, no. 1, p. 1, 2023.

[15] A. E. Johnson, T. J. Pollard, L. Shen, L.-w. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, and R. G. Mark, "Mimic-iii, a freely accessible critical care database," *Scientific data*, vol. 3, no. 1, pp. 1–9, 2016.

[16] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[17] B. Li, Y. Luo, Z. Liu, J. Zheng, J. Lv, and Q. Ma, "Hyperimts: Hypergraph neural network for irregular multivariate time series forecasting," *arXiv preprint arXiv:2505.17431*, 2025.

[18] Y. Luo, B. Zhang, Z. Liu, and Q. Ma, "Hi-patch: Hierarchical patch gnn for irregular multivariate time series," in *Forty-second International Conference on Machine Learning*.

[19] X. Zhang, M. Zeman, T. Tsiligkaridis, and M. Zitnik, "Graph-guided network for irregularly sampled multivariate time series," *arXiv preprint arXiv:2110.05357*, 2021.

[20] H. Wang, J. Peng, F. Huang, J. Wang, J. Chen, and Y. Xiao, "Micn: Multi-scale local and global context modeling for long-term series forecasting," in *The eleventh international conference on learning representations*, 2023.

[21] Z. Shang, L. Chen, B. Wu, and D. Cui, "Ada-mshyper: adaptive multi-scale hypergraph transformer for time series forecasting," *Advances in Neural Information Processing Systems*, vol. 37, pp. 33310–33337, 2024.

[22] A. Das, W. Kong, A. Leach, S. Mathur, R. Sen, and R. Yu, "Long-term forecasting with tide: Time-series dense encoder," *arXiv preprint arXiv:2304.08424*, 2023.

[23] S. Wang, H. Wu, X. Shi, T. Hu, H. Luo, L. Ma, J. Y. Zhang, and J. Zhou, "Timemixer: Decomposable multiscale mixing for time series forecasting," *arXiv preprint arXiv:2405.14616*, 2024.

[24] H. Harutyunyan, H. Khachatrian, D. C. Kale, G. Ver Steeg, and A. Galstyan, "Multitask learning and benchmarking with clinical time series data," *Scientific data*, vol. 6, no. 1, p. 96, 2019.