Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

Hybrid Temporal Modeling and Selective Spectral Transport for Time-Series Imputation

Zongye Tao
Dalian Maritime University, China

***Abstract*—Time-series imputation remains a fundamental challenge due to the coexistence of complex temporal dependencies and structured missing patterns. While existing approaches have achieved strong performance in recovering missing values, they are predominantly driven by point-wise reconstruction in the time domain, which often fails to preserve intrinsic temporal structures under severe and non-uniform missingness. In particular, local temporal patterns and long-range dynamics can be significantly distorted, leading to inconsistent or unstable imputations.**

In this paper, we propose a unified framework that addresses time-series imputation from both temporal modeling and structural alignment perspectives. The proposed model jointly captures short-range dependencies and long-range temporal dynamics through a collaborative sequence modeling design, enabling more expressive representation of complex temporal behaviors. To further enforce structural consistency, we introduce a selective spectral transport regularization that aligns the local frequency-domain distributions between imputed and target sequences on missing-intensive regions. This mechanism provides an explicit constraint on temporal patterns beyond point-wise value recovery, encouraging the model to produce structurally faithful imputations.

Extensive experiments demonstrate that the proposed method achieves consistent improvements over strong baselines across diverse datasets and missing settings, particularly under structured and high missing-rate scenarios.

***Index Terms*—time-series imputation, structured missingness, spectral transport, temporal modeling, long-short dependency modeling**

I. Introduction

MULTIVARIATE time-series imputation is a fundamental problem in data-driven applications, since missing observations can substantially impair the reliability of downstream analysis and decision making [1]. In real-world scenarios, missing values rarely occur as independent random noise. Instead, they often appear in contiguous intervals or irregular blocks due to device failures, communication interruptions, and imperfect collection processes [2]. Such structured missingness breaks temporal continuity, weakens observable cross-variable interactions, and makes it difficult to recover the underlying dynamics of the original sequence.

Recent advances in deep learning have significantly improved time-series imputation by leveraging expressive sequence models [3]–[5].. Existing approaches mainly estimate the missing entries from observed context through predictive or generative modeling, and have shown promising performance on standard benchmarks [3]–[5]. However, most of them are still dominated by point-wise reconstruction in the time domain. Although such objectives can improve numerical recovery, they do not explicitly constrain whether the imputed sequence preserves intrinsic temporal structures, such as local patterns, periodic behaviors, and long-range dynamics. This often leads to over-smoothed or structurally inconsistent imputations under severe or structurally biased missingness, as illustrated in Fig. 1. In particular, although point-wise reconstruction can produce numerically plausible values, it fails to preserve intrinsic temporal structures, resulting in inconsistencies that become more evident when viewed in the frequency domain. Consequently, the recovered sequence may appear accurate in value space while deviating from the true temporal dynamics.

This limitation is particularly critical for multivariate time series, where accurate imputation requires the model to simultaneously address two coupled challenges. First, the temporal evolution of a sequence contains both short-range dependencies and long-range dynamics, which are often difficult to capture well using a single modeling mechanism. Second, the quality of imputation should not be measured only by point-wise errors, because two sequences with similar values may still exhibit substantially different temporal patterns in the structural sense. Therefore, a robust imputation model should not only recover missing values from observed context, but also preserve the temporal structures that characterize the original sequence.

Motivated by these observations, we propose a unified framework for time-series imputation that addresses the problem from both temporal modeling and structural alignment perspectives. On the temporal side, the proposed framework adopts a long-short collaborative design to jointly capture local dependencies and long-range dynamics, enabling expressive representations for complex multivariate sequences. On the structural side, we introduce a selective spectral transport regularization that aligns local frequency-domain distributions between imputed and target sequences on missing-intensive regions. By enforcing structural consistency beyond point-wise reconstruction, the proposed method encourages the model to generate imputations that are not only accurate in value space but also faithful in temporal pattern space.

The main contributions of this paper are summarized as follows:

- We propose a unified framework for multivariate time-series imputation that jointly models temporal dependency learning and structural consistency preservation under structured missingness.
- We develop a long-short collaborative sequence modeling strategy that improves the representation of both local temporal patterns and long-range dynamics for imputation.
- We introduce a selective spectral transport regularization that explicitly aligns local temporal structures in the fre-

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

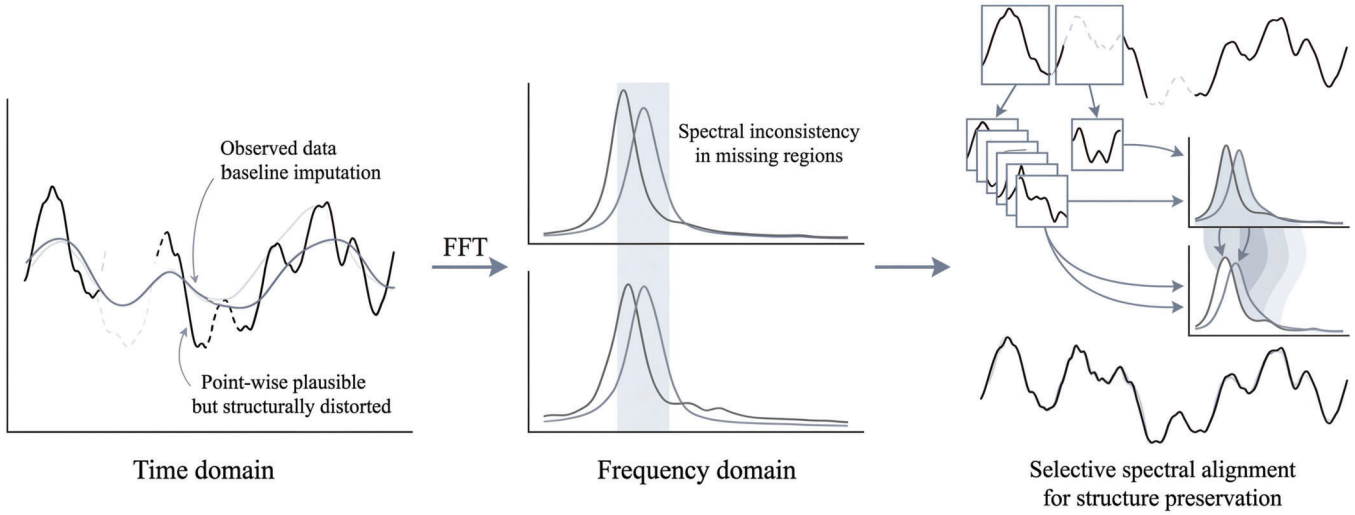

Fig. 1. Illustration of the limitation of point-wise time-series imputation and the proposed selective spectral alignment. Left: point-wise imputation produces numerically plausible but structurally distorted results in the time domain. Middle: such discrepancies become more evident in the frequency domain, especially in missing regions. Right: the proposed method performs selective spectral alignment on missing-intensive regions to preserve temporal structures.

quency domain, providing a structural constraint beyond conventional point-wise recovery.

- Extensive experiments on real-world datasets demonstrate that the proposed method consistently improves imputation performance and shows strong robustness under diverse missing settings.

II. Related Work

A. Time-Series Imputation

Time-series imputation has been extensively studied in both statistical and deep learning paradigms. Early approaches, including mean imputation, k-nearest neighbors, and regression-based methods, are computationally efficient but limited in capturing complex temporal dependencies. Recent deep learning methods leverage neural sequence models to improve imputation performance by exploiting temporal context and cross-variable interactions. These approaches typically formulate imputation as a prediction or reconstruction task, estimating missing values from observed data. Despite their effectiveness, most existing methods are primarily driven by point-wise reconstruction objectives in the time domain, which may lead to numerically plausible results while failing to preserve intrinsic temporal structures under structured or high missingness scenarios.

B. Temporal Dependency Modeling

Modeling temporal dependencies is central to time-series analysis. Recurrent neural networks and their variants have been widely used to capture sequential dynamics [3], [4], but they often suffer from limited efficiency and difficulties in modeling long-range dependencies. More recently, attention-based architectures have demonstrated strong capability in learning long-range interactions and enabling parallel computation [6]. However, these methods typically rely on a unified modeling mechanism, which may not effectively capture the coexistence of short-range patterns and long-range temporal dynamics. This limitation suggests the need for collaborative modeling strategies that can jointly represent multiple scales of temporal dependencies [7], [8].

C. Structure-Aware and Distribution-Based Modeling

Beyond temporal modeling, recent studies have explored incorporating structural or distributional information into sequence learning. In particular, optimal transport has been introduced as a principled way to measure discrepancies between distributions and has been applied in various sequence modeling tasks [9], [10]. Additionally, frequency-domain analysis provides an alternative perspective for capturing temporal patterns, such as periodicity and trend behaviors [11], [12]. However, these approaches are not specifically designed for time-series imputation under structured missingness, and they often lack mechanisms to selectively enforce structural consistency in missing-intensive regions. This motivates the development of methods that integrate temporal modeling with structure-aware alignment in a unified framework.

III. Method

A. Problem Formulation

Let $\mathbf{X} \in \mathbb{R}^{T\times D}$ denote a multivariate time series with T time steps and D variables, and let $\mathbf{M} \in \{0,1\}^{T\times D}$ denote the observation mask, where $M_{t,d} = 1$ indicates that $X_{t,d}$ is observed and $M_{t,d} = 0$ otherwise. The partially observed sequence is written as

$$\mathbf{X}_{\text{obs}} = \mathbf{M} \odot \mathbf{X}. \tag{1}$$

The goal of time-series imputation is to learn a mapping

$$\hat{\mathbf{X}} = f(\mathbf{X}_{\text{obs}}, \mathbf{M}), \tag{2}$$

such that $\hat{\mathbf{X}}$ approximates the complete sequence $\mathbf{X}$. Unlike conventional formulations that rely solely on point-wise recovery, our objective is to jointly preserve temporal dependency and structural consistency.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

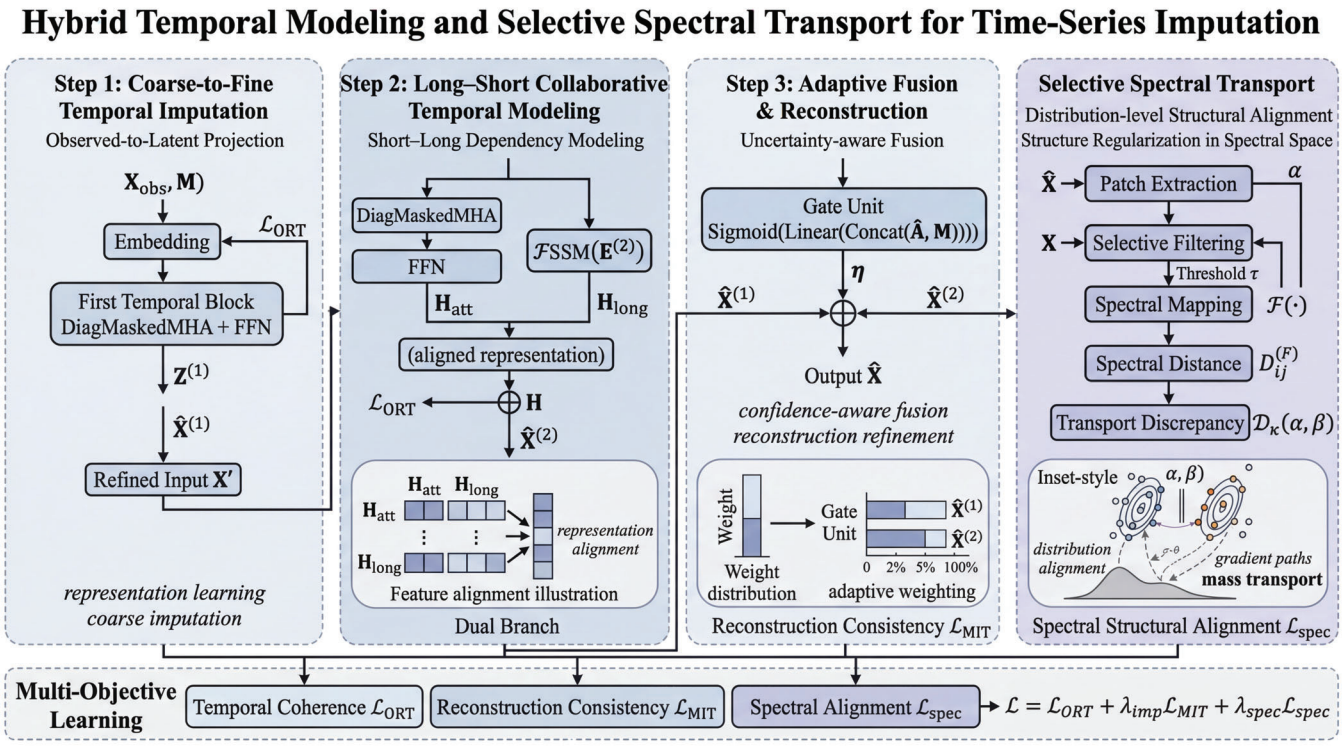

Fig. 2. Overall framework of the proposed hybrid temporal modeling and selective spectral transport method for time-series imputation. The framework integrates coarse-to-fine temporal imputation, long-short collaborative modeling, adaptive fusion, and selective spectral transport to jointly capture multi-scale temporal dependencies and enforce structural consistency in missing-intensive regions under a unified optimization objective.

B. Overall Framework

As illustrated in Fig. 2, the proposed method is a unified framework that combines temporal modeling and spectral structural alignment for time-series imputation. The framework follows a two-stage formulation. The first stage performs temporal imputation through a dual-block architecture, where the incomplete sequence and the missing mask are jointly encoded to produce progressive representations for coarse-to-fine recovery. The second stage introduces structure-aware refinement by aligning local spectral distributions between imputed and target segments in missing-intensive regions, thereby encouraging structurally consistent reconstruction beyond point-wise recovery. Specifically, the framework consists of coarse-to-fine temporal imputation, long-short collaborative temporal modeling, adaptive fusion and reconstruction, and selective spectral transport regularization under a joint multi-objective learning scheme. Formally, let $\hat{\mathbf{X}}^{(1)}$, $\hat{\mathbf{X}}^{(2)}$, and $\hat{\mathbf{X}}^{(3)}$ denote the outputs from the first temporal block, the long-short refinement block, and the final fusion block, respectively. The complete imputed sequence is obtained by replacing missing entries in the input with the final learned representation:

$$\hat{\mathbf{X}} = \mathbf{M} \odot \mathbf{X}_{\text{obs}} + (1 - \mathbf{M}) \odot \hat{\mathbf{X}}^{(3)}. \tag{3}$$

This design follows a staged reconstruction strategy, where coarse estimation, refinement, and adaptive fusion are performed sequentially. While the temporal module focuses on dependency learning in the time domain, the spectral transport module enforces structural consistency at the distribution level.

C. Long-Short Collaborative Temporal Modeling

1) Embedding and Dual-Branch Temporal Encoding: Given the incomplete input, we first concatenate the feature values and the missing mask, and project them into a latent space:

$$\mathbf{E}^{(1)} = \text{Linear}(\text{Concat}(\mathbf{X}_{\text{obs}}, \mathbf{M})) + \mathbf{P}, \tag{4}$$

where $\mathbf{P}$ denotes positional encoding. This formulation jointly encodes observed values and missing patterns before temporal modeling, enabling the model to exploit both content information and missingness cues.

To capture short-range temporal interactions, we use a diagonally-masked self-attention block. Given query, key, and value matrices $\mathbf{Q}, \mathbf{K}, \mathbf{V}$, the standard self-attention is

$$\text{Attn}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}}\right)\mathbf{V}. \tag{5}$$

Following the diagonally-masked design, the diagonal entries of the attention logits are suppressed so that each time step is estimated from other steps rather than trivially copying itself:

$$\text{DiagMask}(\mathbf{A})_{ij} = \begin{cases} -\infty, & i = j, \\ A_{ij}, & i \neq j, \end{cases} \tag{6}$$

and the resulting diagonally-masked self-attention becomes

$$\text{DMSA}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\left(\text{DiagMask}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}}\right)\right)\mathbf{V}. \tag{7}$$

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

This mechanism prevents trivial self-copying and enforces dependency learning across time steps.

Using multi-head diagonal masking and a feed-forward network, the first temporal block is written as

$$\mathbf{Z}^{(1)} = \left\{\mathrm{FFN}\left(\mathrm{DiagMaskedMHA}\left(\mathbf{E}^{(1)}\right)\right)\right\}^{N}, \tag{8}$$

$$\hat{\mathbf{X}}^{(1)} = \mathrm{Linear}(\mathbf{Z}^{(1)}). \tag{9}$$

We then replace missing positions in the input with this first-stage estimate:

$$\mathbf{X}' = \mathbf{M} \odot \mathbf{X}_{\mathrm{obs}} + (1 - \mathbf{M}) \odot \hat{\mathbf{X}}^{(1)}. \tag{10}$$

Next, a second temporal block further refines the sequence:

$$\mathbf{E}^{(2)} = \mathrm{Linear}(\mathrm{Concat}(\mathbf{X}', \mathbf{M})) + \mathbf{P}, \tag{11}$$

$$\mathbf{Z}^{(2)} = \left\{\mathrm{FFN}\left(\mathrm{DiagMaskedMHA}\left(\mathbf{E}^{(2)}\right)\right)\right\}^{N}, \tag{12}$$

$$\mathbf{H}_{\mathrm{att}} = \mathrm{Linear}\left(\mathrm{ReLU}\left(\mathrm{Linear}(\mathbf{Z}^{(2)})\right)\right). \tag{13}$$

This two-stage design follows a progressive refinement strategy for sequential imputation, where the second block acts on the first imputed result instead of the raw incomplete input.

2) Long-Range Dynamics via Selective State Space Modeling: While DMSA is effective at extracting short-range temporal relations and cross-step interactions, long-range dependency is additionally modeled through a selective state-space branch.

A continuous-time state-space system is defined as

$$\mathbf{h}'(t) = A\mathbf{h}(t) + B\mathbf{x}(t), \qquad \mathbf{y}(t) = C\mathbf{h}(t), \tag{14}$$

where $\mathbf{h}(t) \in \mathbb{R}^N$ is the latent state and A, B, C are learnable matrices. After discretization with step size Δ, it becomes

$$\mathbf{h}_t = \bar{A}\mathbf{h}_{t-1} + \bar{B}\mathbf{x}_t, \qquad \mathbf{y}_t = C\mathbf{h}_t, \tag{15}$$

where $\bar{A}$ and $\bar{B}$ are the discretized transition matrices. Equivalently, the same model admits a convolutional form

$$\mathbf{K} = (C\bar{B}, C\bar{A}\bar{B}, \ldots, C\bar{A}^k\bar{B}, \ldots), \qquad \mathbf{y} = \mathbf{x} * \mathbf{K}, \tag{16}$$

which provides an efficient view of long-range sequence propagation.

Based on this formulation, the long-range branch is written as

$$\mathbf{H}_{\mathrm{long}} = \mathcal{F}_{\mathrm{SSM}}(\mathbf{E}^{(2)}), \tag{17}$$

while the short-range branch from the second temporal block is denoted as

$$\mathbf{H}_{\mathrm{short}} = \mathbf{H}_{\mathrm{att}}. \tag{18}$$

We combine the two representations to obtain the final temporal representation

$$\mathbf{H} = \mathbf{H}_{\mathrm{short}} + \mathbf{H}_{\mathrm{long}}, \tag{19}$$

followed by a projection head

$$\hat{\mathbf{X}}^{(2)} = \mathrm{Linear}(\mathbf{H}). \tag{20}$$

This collaborative formulation keeps the short-range discrimination of attention while injecting long-range propagation through the selective state-space dynamics.

3) Adaptive Fusion of Multi-Stage Representations: To combine the coarse and refined estimates adaptively, we compute fusion weights from the attention map and the missing mask, following a weighted fusion strategy based on attention and missing patterns. Let $\hat{\mathbf{A}}$ be the averaged attention weights from the last attention layer:

$$\hat{\mathbf{A}} = \frac{1}{h}\sum_{i=1}^{h} \mathbf{A}_i. \tag{21}$$

Then the fusion gate is defined as

$$\boldsymbol{\eta} = \mathrm{Sigmoid}\left(\mathrm{Linear}\left(\mathrm{Concat}(\hat{\mathbf{A}}, \mathbf{M})\right)\right), \tag{22}$$

and the final learned representation is

$$\hat{\mathbf{X}}^{(3)} = (1 - \boldsymbol{\eta}) \odot \hat{\mathbf{X}}^{(1)} + \boldsymbol{\eta} \odot \hat{\mathbf{X}}^{(2)}. \tag{23}$$

Finally,

$$\hat{\mathbf{X}} = \mathbf{M} \odot \mathbf{X}_{\mathrm{obs}} + (1 - \mathbf{M}) \odot \hat{\mathbf{X}}^{(3)}. \tag{24}$$

D. Selective Spectral Transport Regularization

Point-wise reconstruction alone does not explicitly enforce temporal pattern consistency. As illustrated in Fig. 3, conventional imputation methods may produce values that are numerically plausible while still distorting the underlying temporal structures in missing regions. This discrepancy becomes more evident in the frequency domain, where key spectral components can be severely misaligned. To address this issue, we introduce a selective spectral transport regularizer defined on local windows extracted from missing-intensive regions.

1) Patch Extraction and Spectral Distance: We first sample two sets of temporal patches, denoted by

$$\alpha = \{\alpha_i\}_{i=1}^{n}, \qquad \beta = \{\beta_j\}_{j=1}^{m}, \tag{25}$$

from the imputed sequence and the corresponding target sequence, respectively, where α and β denote all sampled patches before selective filtering. Unlike standard Wasserstein discrepancy, which uses Euclidean distances in the time domain,

$$W(\alpha, \beta) = \min_{T \in \Pi(\alpha,\beta)} \langle D, T \rangle, \tag{26}$$

where $\Pi(\alpha, \beta)$ denotes the set of valid transport plans between distributions α and β, and $D_{ij} = \|\alpha_i - \beta_j\|_2$. We replace this pairwise distance with a frequency-domain comparison that captures temporal patterns.

For each patch, we apply the discrete Fourier transform:

$$\mathcal{F}(\alpha_i) \quad \text{and} \quad \mathcal{F}(\beta_j) \tag{27}$$

and define the pairwise spectral distance as

$$D_{ij}^{(F)} = \left\|\mathcal{F}(\alpha_i) - \mathcal{F}(\beta_j)\right\|_1. \tag{28}$$

This formulation enables the model to encode temporal patterns through frequency-domain representations, which are more informative for capturing periodicity and dynamic structures.

The corresponding spectral Wasserstein discrepancy is

$$W_{\mathrm{spec}}(\alpha, \beta) = \min_{T \in \Pi(\alpha,\beta)} \langle D^{(F)}, T \rangle. \tag{29}$$

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

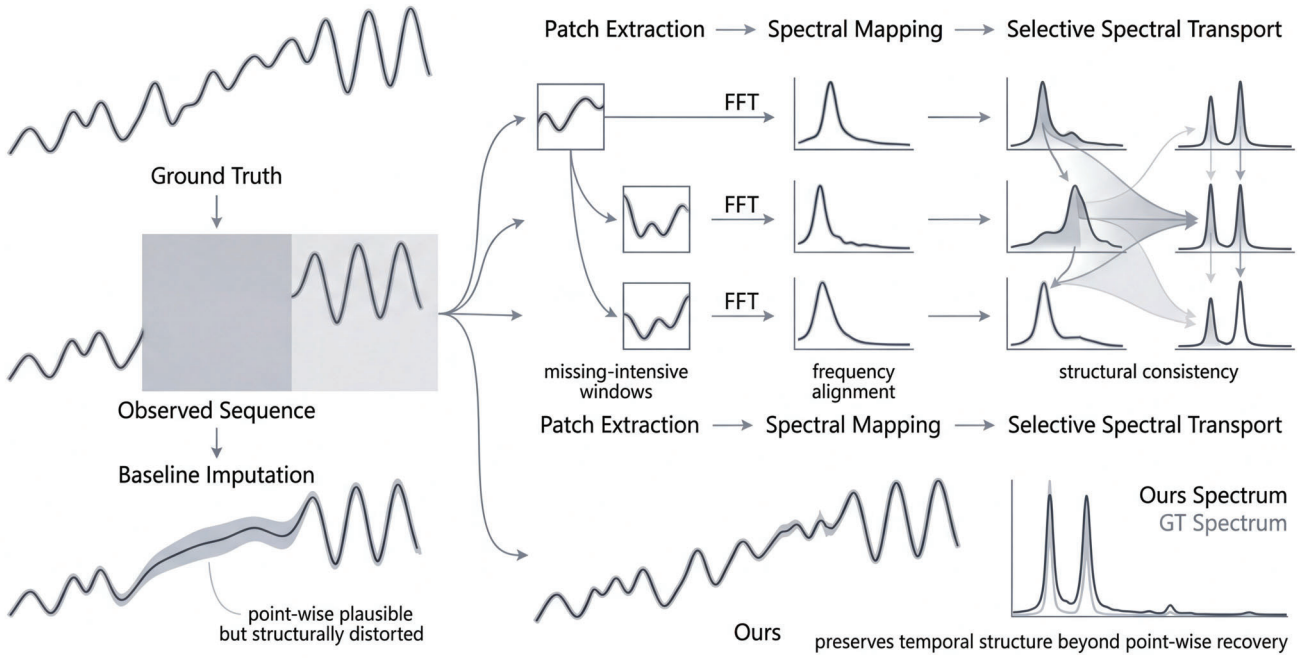

Fig. 3. Illustration of spectral structural inconsistency in time-series imputation and the proposed selective spectral transport mechanism. Left: In the time domain, baseline imputation produces values that are numerically plausible but structurally distorted in missing regions. Middle: The proposed method extracts missing-intensive patches and maps them into the frequency domain via FFT, where spectral discrepancies are identified. Right: Through selective spectral transport, the frequency components are aligned with those of the ground truth, resulting in structurally consistent reconstruction beyond point-wise recovery.

2) Selective Matching Regularization: To account for non-stationarity and avoid forcing all patches to match, we adopt the regularized spectral transport discrepancy:

$$\mathcal{D}_\kappa(\alpha,\beta) = \min_{T\geq 0}\langle D^{(F)}, T\rangle + \kappa\Big(D_{\mathrm{KL}}(T\mathbf{1}_m\|\Delta_n) + D_{\mathrm{KL}}(T^\top\mathbf{1}_n\|\Delta_m)\Big) \tag{30}$$

where $\Delta_n = \mathbf{1}_n/n$ and $\Delta_m = \mathbf{1}_m/m$ are uniform simplex vectors, and κ controls the matching strength. This regularization improves robustness compared to standard transport formulations under non-stationary temporal patterns.

In our imputation setting, we do not apply this discrepancy to all windows uniformly. Instead, we only retain windows whose missing ratio exceeds a threshold τ, yielding the selective patch sets

$$\alpha_{\mathrm{sel}} = \{\alpha_i : r_i \geq \tau\}, \qquad \beta_{\mathrm{sel}} = \{\beta_j : r_j \geq \tau\}, \tag{31}$$

where r_i and r_j denote the missing ratios of the corresponding windows, and α_{sel} and β_{sel} are the filtered subsets used for selective alignment. The structural regularization term is then defined as

$$\mathcal{L}_{\mathrm{spec}} = \mathcal{D}_\kappa(\alpha_{\mathrm{sel}}, \beta_{\mathrm{sel}}). \tag{32}$$

E. Joint Optimization Objective

Following a joint optimization strategy with both observed and masked supervision, we use both masked imputation and observed reconstruction during training rather than relying on a single reconstruction target. Let $\hat{\mathbf{M}}$ denote the observation mask after artificial masking and let $\mathbf{I}$ denote the indicator of artificially masked entries. The masked imputation loss is

$$\ell_{\mathrm{MAE}}(\mathbf{A},\mathbf{B},\mathbf{S}) = \frac{\sum_{t=1}^{T}\sum_{d=1}^{D}|(\mathbf{A}-\mathbf{B})\odot\mathbf{S}|_{t,d}}{\sum_{t=1}^{T}\sum_{d=1}^{D}S_{t,d}}, \tag{33}$$

$$\mathcal{L}_{\mathrm{MIT}} = \ell_{\mathrm{MAE}}(\hat{\mathbf{X}}, \mathbf{X}, \mathbf{I}), \tag{34}$$

and the observed reconstruction loss is

$$\mathcal{L}_{\mathrm{ORT}} = \frac{1}{3}\Big(\ell_{\mathrm{MAE}}(\hat{\mathbf{X}}^{(1)}, \mathbf{X}, \hat{\mathbf{M}}) + \ell_{\mathrm{MAE}}(\hat{\mathbf{X}}^{(2)}, \mathbf{X}, \hat{\mathbf{M}}) + \ell_{\mathrm{MAE}}(\hat{\mathbf{X}}^{(3)}, \mathbf{X}, \hat{\mathbf{M}})\Big) \tag{35}$$

which applies multi-stage supervision by involving all intermediate learned representations in reconstruction loss computation.

Finally, the overall training objective is

$$\mathcal{L} = \mathcal{L}_{\mathrm{ORT}} + \lambda_{\mathrm{imp}}\mathcal{L}_{\mathrm{MIT}} + \lambda_{\mathrm{spec}}\mathcal{L}_{\mathrm{spec}}. \tag{36}$$

Here, $\mathcal{L}_{\mathrm{ORT}}$ preserves fidelity on observed entries, $\mathcal{L}_{\mathrm{MIT}}$ explicitly supervises missing-value recovery, and $\mathcal{L}_{\mathrm{spec}}$ regularizes the structural consistency of imputed temporal patterns in the frequency domain.

IV. Experiments and Results

A. Experimental Settings

Datasets: We evaluate the proposed method on four widely used real-world multivariate time-series datasets, covering healthcare, environmental monitoring, energy consumption,

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

TABLE I. General statistics of the datasets used in our experiments.

	PhysioNet-2012	Air-Quality	Electricity	ETTm1
Number of samples	11,988	1,461	1,400	69,680
Number of variables	37	132	370	7
Sequence length	48	24	100	24
Original missing rate	80.67%	1.6%	0%	0%

and industrial sensing scenarios. These datasets exhibit diverse temporal characteristics and missing patterns, which provide a comprehensive benchmark for time-series imputation.

- **PhysioNet-2012:** This dataset contains multivariate clinical records collected from intensive care unit patients during the first 48 hours after admission [13].Each sample consists of 37 physiological variables, and the data are highly sparse due to irregular measurement frequencies and incomplete observations. Following common practice, we split the dataset into training, validation, and test sets with a ratio of 64%/16%/20%. To evaluate imputation quality, we randomly mask 10% of the observed entries in the validation and test sets and use them as ground truth.
- **Air-Quality:** The Beijing Multi-Site Air-Quality dataset includes hourly pollutant measurements collected from 12 monitoring stations [14]. We aggregate all station-wise variables into a 132-dimensional multivariate sequence. The dataset covers 48 months, and each sample is formed by 24 consecutive time steps. We use the earliest 10 months for testing, the next 10 months for validation, and the remaining period for training. As in PhysioNet-2012, 10% of observed values in the validation and test sets are additionally masked for evaluation.
- **Electricity:** This dataset records electricity consumption from 370 clients at 15-minute intervals [15]. Since the raw dataset is complete, we construct imputation benchmarks by introducing artificial missingness. Each sample contains 100 consecutive time steps. We partition the data chronologically into training, validation, and test subsets using the same protocol as in prior imputation studies, and randomly hold out observed entries for evaluation.
- **ETTm1:** ETTm1 is a widely used benchmark derived from electricity transformer temperature monitoring [16]. It contains 7 variables that reflect transformer operating conditions and external load information. We follow the standard chronological split and use sliding windows of length 24 to generate samples. Similar to Electricity, the original data are complete, and artificial missing values are introduced in the validation and test sets for evaluation.

Baselines: We compare our method with representative time-series imputation approaches from different categories, including two simple statistical methods (*Median* and *Last*), recurrent models (*M-RNN* [3] and *BRITS* [4]), probabilistic and generative methods (*GP-VAE* [5] and *CSDI* [17]), and attention-based models (*Transformer* [6] and *SAITS* [18]). These baselines cover both classical and recent deep imputation paradigms, including recurrent, adversarial, probabilistic, and attention-based methods [19], [20], allowing a comprehensive assessment of the proposed framework.

Implementation Details: All experiments are implemented in PyTorch and conducted on NVIDIA GPUs. For fair comparison, we use the same data split and evaluation protocol across all methods. The proposed model is optimized with Adam, and the best checkpoint is selected according to validation MAE. For datasets without natural missing values, we simulate missingness by randomly masking observed entries following standard benchmark settings.

Model Settings: The temporal module uses stacked diagonal-masked attention blocks together with a state-space branch for long-range modeling [7], [8]. The spectral alignment module is activated only on selected windows whose missing ratios exceed a predefined threshold. Unless otherwise stated, the hyperparameters are tuned on the validation set. The key hyperparameters include the hidden dimension, number of attention heads, number of temporal blocks, learning rate, batch size, spectral window size, stride, matching strength, and the weight of the structural regularization term.

Evaluation Metrics: We report Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Relative Error (MRE), all computed only on the held-out missing entries. Given the ground-truth sequence $\mathbf{X}$, the imputed sequence $\hat{\mathbf{X}}$, and the evaluation mask $\mathbf{M}$, the metrics are defined as follows:

$$\mathrm{MAE}(\mathbf{X},\widetilde{\mathbf{X}},\mathbf{M}) = \frac{\sum_{t_i=t_1}^{t_n}\sum_{j=1}^{d}\left|\left(x_{t_i}^{j}-\widetilde{x}_{t_i}^{j}\right)\cdot m_{t_i}^{j}\right|}{\sum_{t_i=t_1}^{t_n}\sum_{j=1}^{d} m_{t_i}^{j}} \tag{37}$$

$$\mathrm{RMSE}(\mathbf{X},\widetilde{\mathbf{X}},\mathbf{M}) = \sqrt{\frac{\sum_{t_i=t_1}^{t_n}\sum_{j=1}^{d}\left[\left(x_{t_i}^{j}-\widetilde{x}_{t_i}^{j}\right)^{2}\cdot m_{t_i}^{j}\right]}{\sum_{t_i=t_1}^{t_n}\sum_{j=1}^{d} m_{t_i}^{j}}} \tag{38}$$

$$\mathrm{MRE}(\mathbf{X},\widetilde{\mathbf{X}},\mathbf{M}) = \frac{\sum_{t_i=t_1}^{t_n}\sum_{j=1}^{d}\left|\left(x_{t_i}^{j}-\widetilde{x}_{t_i}^{j}\right)\cdot m_{t_i}^{j}\right|}{\sum_{t_i=t_1}^{t_n}\sum_{j=1}^{d}\left|\widetilde{x}_{t_i}^{j}\cdot m_{t_i}^{j}\right|} \tag{39}$$

Lower values indicate better imputation performance.

B. Overall Imputation Performance

Table III reports the imputation performance of all compared methods on the four benchmark datasets. Overall, the proposed method achieves the best or highly competitive results across all datasets and evaluation metrics, demonstrating its effectiveness under different temporal characteristics, feature dimensions, and missing patterns.

On **PhysioNet-2012**, our method achieves the best MAE of 0.188 and the best MRE of 27.0%, outperforming both recurrent and attention-based baselines. Compared with SAITS, which is the strongest baseline on this dataset, our method

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

TABLE II. Main hyperparameter settings of the proposed method.

	PhysioNet-2012	Air-Quality	Electricity	ETTm1
Batch size	128	128	128	128
Learning rate	0.0006828	0.0008821	0.0003592	0.0004290
d_{model}	256	512	1024	1024
Number of heads	8	4	8	8
Number of temporal blocks	5	1	1	1
Dropout	0	0	0.2	0.1
Use Mamba	True	True	True	True
Mamba state dim	4	4	4	4
Mamba expansion	2	2	2	2

TABLE III. Performance comparison for Imputation

Method	PhysioNet-2012	Air-Quality	Electricity	ETT(m1)
Median	0.725 / 0.985 / 103.1%	0.761 / 1.173 / 107.2%	2.088 / 2.739 / 112.1%	1.151 / 1.862 / 139.9%
Last	0.850 / 1.188 / 118.3%	0.948 / 1.389 / 129.1%	0.988 / 1.497 / 51.3%	0.977 / 1.257 / 95.9%
GRUI-GAN	0.760 / 1.021 / 107.0%	0.766 / 1.155 / 108.1%	/	0.603 / 0.710 / 94.2%
E^2GAN	0.698 / 0.944 / 99.1%	0.752 / 1.124 / 105.7%	/	0.558 / 0.688 / 87.0%
M-RNN	0.521 / 0.765 / 74.9%	0.302 / 0.651 / 42.0%	1.231 / 1.848 / 65.8%	0.389 / 0.446 / 32.1%
GP-VAE	0.410 / 0.655 / 57.1%	0.273 / 0.632 / 38.3%	1.113 / 1.591 / 58.9%	0.277 / 0.322 / 15.9%
BRITS	0.259 / 0.780 / 36.5%	0.159 / 0.554 / 22.6%	0.881 / 1.333 / 46.3%	0.133 / 0.264 / 12.8%
Transformer	0.197 / 0.466 / 27.5%	0.168 / 0.571 / 23.3%	0.877 / 1.327 / 45.6%	0.121 / 0.193 / 11.8%
SAITS	0.192 / 0.439 / 27.3%	0.146 / 0.521 / 20.6%	0.822 / 1.221 / 44.0%	0.121 / 0.197 / 11.6%
Ours	0.188 / 0.457 / 27.0%	0.130 / 0.331 / 18.4%	0.807 / 1.136 / 43.2%	0.097 / 0.163 / 9.3%

further reduces MAE from 0.192 to 0.188 and MRE from 27.3% to 27.0%. Although the RMSE of our method (0.457) is slightly higher than that of SAITS (0.439), the overall improvements in MAE and MRE indicate that the proposed framework provides more stable point-wise recovery under highly sparse and irregular clinical observations. Since PhysioNet-2012 has an original missing rate of 80.67%, this result suggests that the proposed long-short temporal modeling can effectively exploit limited observations, while the spectral regularization helps preserve local temporal consistency in severely incomplete sequences.

On the **Air-Quality** dataset, our method shows a clear advantage over all baselines, achieving 0.130 MAE, 0.331 RMSE, and 18.4% MRE. Compared with SAITS, the strongest competing method, our model reduces MAE from 0.146 to 0.130 and RMSE from 0.521 to 0.331. The large RMSE reduction is particularly noteworthy, as it indicates that our method is more effective at suppressing large reconstruction deviations. This improvement can be attributed to the fact that air-quality data often contain strong local periodicity and station-related structural correlations. By introducing selective spectral transport on missing-intensive windows, the proposed method can better preserve latent frequency-domain patterns beyond conventional time-domain reconstruction.

On the high-dimensional **Electricity** dataset, our method again achieves the best results, with MAE/RMSE/MRE of 0.807/1.136/43.2%. Compared with SAITS, the improvement is consistent across all three metrics. Although the margins are smaller than those on Air-Quality and ETTm1, this result remains important because Electricity contains 370 variables and relatively long observation windows, making representation learning considerably more challenging. The performance gain demonstrates that the proposed framework scales well to high-dimensional multivariate sequences and remains robust even when the original data do not contain natural missingness.

On **ETTm1**, our method achieves the best performance by a clear margin, with MAE 0.097, RMSE 0.163, and MRE 9.3%. Compared with Transformer and SAITS, which already provide strong performance on this dataset, our model further reduces the MAE from 0.121 to 0.097. This substantial improvement indicates that the collaborative design between diagonal-masked attention and the state-space branch is particularly beneficial for data with relatively regular temporal dynamics and long-range dependency patterns. The result also suggests that the proposed method is not only effective on sparse real-world data, but also highly competitive on structured industrial time-series benchmarks.

From a broader perspective, several trends can be observed from Table III. First, simple statistical methods such as *Median* and *Last* perform poorly on most datasets, especially under complex temporal dynamics, confirming that naive interpolation strategies are insufficient for realistic multivariate imputation. Second, recurrent and generative methods such as M-RNN, BRITS, and GAN-based imputation models are more effective than statistical baselines, but still fall behind recent attention-based models on most benchmarks [3], [4], [19]. Third, attention-based methods, especially Transformer and SAITS, achieve strong results due to their ability to

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

TABLE IV. The results of the downstream classification task on the PhysioNet-2012 dataset are presented. Performance metrics for each method are based on five independent runs, with the reported values representing the means ± standard deviations. Higher values indicate better performance.

Method	ROC-AUC	PR-AUC	F1-score
Median	83.4% ± 0.5%	46.1% ± 0.6%	38.4% ± 3.0%
Last	82.5% ± 0.4%	46.8% ± 0.6%	39.2% ± 2.3%
GRUI-GAN	83.0% ± 0.4%	45.2% ± 0.7%	38.6% ± 2.2%
E^2GAN	82.9% ± 0.3%	45.1% ± 0.8%	36.2% ± 2.3%
M-RNN	82.5% ± 0.3%	45.4% ± 0.6%	38.4% ± 3.1%
GP-VAE	83.6% ± 0.3%	48.0% ± 0.9%	40.7% ± 3.5%
BRITS	83.7% ± 0.2%	49.1% ± 0.6%	41.7% ± 1.7%
Transformer	84.1% ± 0.8%	49.2% ± 1.6%	41.4% ± 2.1%
SAITS	84.7% ± 0.5%	51.0% ± 0.8%	42.7% ± 2.9%
Ours	**86.1% ± 0.3%**	**55.4% ± 0.5%**	**52.8% ± 2.1%**

capture cross-time interactions. However, the proposed method consistently improves upon these baselines, which validates the importance of combining temporal dependency learning with explicit structural alignment.

In summary, the results in Table III demonstrate that the proposed method achieves strong generalization across different application domains, ranging from healthcare and environmental monitoring to energy consumption and industrial sensing. Such consistent gains indicate that the model is able to recover missing values accurately while also preserving intrinsic temporal structures.

C. Downstream Classification Performance

To further evaluate whether the imputed data are useful for practical applications, we conduct a downstream classification experiment on the PhysioNet-2012 dataset. The results are summarized in Table IV.

Our method achieves the best performance on all three metrics, yielding a ROC-AUC of 86.0%, a PR-AUC of 55.3%, and an F1-score of 52.7%. Compared with the strongest baseline SAITS, our method improves ROC-AUC from 84.6% to 86.0%, PR-AUC from 50.9% to 55.3%, and F1-score from 42.6% to 52.7%. The improvement in PR-AUC and F1-score is particularly substantial, indicating that the representations reconstructed by our method are more informative for identifying positive clinical outcomes.

This observation is important because downstream prediction performance depends not only on point-wise imputation accuracy, but also on whether the recovered sequence preserves discriminative temporal patterns. A model may produce numerically plausible imputations while still distorting clinically meaningful dynamics. The superior downstream performance of our method suggests that the proposed framework better preserves such latent temporal structures. In particular, the spectral alignment module appears to provide an additional structural constraint that is beneficial for maintaining task-relevant patterns, rather than merely minimizing reconstruction error.

Another notable observation is that the gap between imputation metrics and downstream task performance is not always perfectly aligned across methods. For example, some baseline methods achieve competitive ROC-AUC values but lag behind more clearly on PR-AUC and F1-score. This indicates that small differences in imputation quality may translate into much larger differences in downstream decision quality, especially in imbalanced clinical settings. Therefore, the downstream classification results provide complementary evidence that the proposed method improves not only numerical recovery but also the semantic usefulness of the imputed data.

D. Robustness under Different Missing Rates

To assess the robustness of the proposed method under increasingly severe missing conditions, we further evaluate all methods on the Electricity dataset with missing rates varying from 20% to 90%. The results are presented in Table V.

Overall, our method consistently achieves the best results across all missing rates. When the missing rate is relatively low, such as 20% or 30%, the proposed method already outperforms all competing approaches. For example, at 20% missingness, our method achieves an MAE of 0.742, compared with 0.763 for SAITS and 0.851 for Transformer. This shows that even when sufficient observations are available, the proposed framework can still exploit temporal and structural information more effectively than existing baselines.

As the missing rate increases, the advantage of our method remains stable. At 50% missingness, our model still achieves the best results, with 0.855 MAE, 1.335 RMSE, and 45.5% MRE. Compared with SAITS, our method reduces MAE from 0.876 to 0.855 and RMSE from 1.377 to 1.335. These consistent gains indicate that the model maintains strong reconstruction capability even when half of the entries are missing.

Under more challenging settings, namely 60% to 90% missingness, the superiority of the proposed method becomes even more meaningful. At 90% missingness, our method still obtains the best performance, with MAE 0.920, RMSE 1.335, and MRE 49.0%, outperforming SAITS (0.933/1.354/49.9%) and Transformer (0.942/1.503/50.2%). Although the absolute margins are not extremely large, maintaining the best performance under such extreme sparsity is difficult and highlights the robustness of the proposed design.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

TABLE V. Performance comparison of different methods on the Electricity dataset across varying missing rates from 20% to 90%. Metrics are reported as MAE / RMSE / MRE (lower is better). The best results are highlighted in bold.

Method	20%	30%	40%	50%
Median	2.058 / 2.735 / 110.2%	2.057 / 2.732 / 110.1%	2.062 / 2.738 / 110.4%	2.050 / 2.726 / 109.7%
Last	1.015 / 1.551 / 54.4%	1.018 / 1.560 / 54.5%	1.026 / 1.578 / 54.9%	1.028 / 1.592 / 55.0%
M-RNN	1.242 / 1.853 / 66.5%	1.258 / 1.879 / 67.2%	1.269 / 1.886 / 68.0%	1.288 / 1.903 / 68.7%
GP-VAE	1.118 / 1.495 / 59.7%	1.014 / 1.543 / 56.4%	1.088 / 1.569 / 58.0%	1.087 / 1.565 / 58.4%
BRITS	0.935 / 1.414 / 50.1%	0.940 / 1.431 / 50.2%	0.990 / 1.492 / 54.1%	1.019 / 1.525 / 54.9%
Transformer	0.853 / 1.327 / 45.6%	0.855 / 1.329 / 45.5%	0.880 / 1.392 / 47.0%	0.899 / 1.416 / 48.0%
SAITS	0.765 / 1.189 / 41.0%	0.788 / 1.221 / 42.1%	0.867 / 1.312 / 46.5%	0.874 / 1.375 / 46.7%
Ours	**0.740 / 1.148 / 39.3%**	**0.770 / 1.187 / 41.2%**	**0.843 / 1.273 / 45.3%**	**0.853 / 1.333 / 45.3%**
Method	60%	70%	80%	90%
Median	2.057 / 2.734 / 110.2%	2.052 / 2.728 / 109.8%	2.062 / 2.737 / 110.4%	2.053 / 2.726 / 110.0%
Last	1.044 / 1.619 / 56.0%	1.044 / 1.640 / 55.9%	1.060 / 1.663 / 56.8%	1.069 / 1.690 / 57.1%
M-RNN	1.293 / 1.908 / 69.1%	1.305 / 1.928 / 69.8%	1.318 / 1.959 / 70.4%	1.335 / 1.973 / 71.9%
GP-VAE	1.101 / 1.619 / 58.9%	1.044 / 1.603 / 55.9%	1.066 / 1.627 / 56.8%	1.010 / 1.625 / 54.1%
BRITS	1.101 / 1.604 / 58.9%	1.090 / 1.609 / 58.4%	1.141 / 1.665 / 61.0%	1.168 / 1.711 / 62.6%
Transformer	0.900 / 1.414 / 48.3%	0.906 / 1.430 / 48.5%	0.923 / 1.470 / 49.6%	0.940 / 1.501 / 50.0%
SAITS	0.890 / 1.326 / 47.7%	0.900 / 1.275 / 48.3%	0.906 / 1.325 / 48.4%	0.931 / 1.352 / 49.7%
Ours	**0.873 / 1.298 / 46.8%**	**0.887 / 1.252 / 47.5%**	**0.893 / 1.298 / 47.6%**	**0.918 / 1.333 / 48.8%**

A key phenomenon revealed by Table V is that the degradation of our method is relatively gradual as the missing rate increases. For instance, the MAE of our model changes from 0.742 at 20% missingness to 0.920 at 90% missingness, whereas most baseline methods show a more pronounced deterioration. This suggests that the proposed framework is less sensitive to observation scarcity. We attribute this robustness to two factors. First, the long-short collaborative temporal module allows the model to capture both local interactions and long-range dynamics, which becomes increasingly important when direct observations are limited. Second, the selective spectral transport regularization explicitly constrains structural consistency in missing-intensive windows, helping the model avoid implausible reconstructions under extreme missingness.

These results demonstrate that the proposed method is not only effective under standard benchmark settings but also reliable in more challenging scenarios where missingness is severe. This property is particularly important in real-world applications, where the missing rate can be highly variable and often much higher than that assumed in standard evaluation protocols.

E. Discussion

The above experiments provide several important insights into the behavior of different imputation methods.

First, the overall comparison shows that methods relying mainly on simple point-wise recovery are insufficient for challenging multivariate time-series imputation. Although recurrent and attention-based models substantially outperform statistical baselines, their performance can still be limited when the missing pattern is highly structured or when the temporal dynamics are complex. This supports our motivation that accurate imputation requires not only value recovery but also structural preservation.

Second, the results suggest that explicitly modeling both short-range and long-range temporal dependencies is beneficial across different data domains. The attention branch is effective at capturing local temporal interactions and cross-step dependencies, while the state-space branch provides an efficient mechanism for long-range propagation. Their collaborative combination appears especially helpful on datasets such as ETTm1 and Electricity, where temporal continuity and longer-range patterns play a major role.

Third, the substantial gain on Air-Quality and the consistent improvements under high missing rates indicate that spectral structural alignment contributes meaningfully beyond conventional time-domain supervision. By selectively applying spectral transport regularization to missing-intensive windows, the model is encouraged to preserve local temporal patterns only where structural distortion is most likely to occur. This selective design avoids unnecessary constraints on easier regions while strengthening recovery in the most challenging parts of the sequence.

Finally, the downstream classification results confirm that better imputation should not be judged solely by reconstruction error. The proposed method produces imputations that are more useful for subsequent predictive modeling, indicating that the learned representations preserve discriminative temporal information rather than only minimizing numerical deviations. Taken together, these findings validate the effectiveness of the proposed unified framework and show that jointly modeling temporal dependencies and structural consistency is a promising direction for robust time-series imputation.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

V. Conclusion

This paper presented a unified framework for multivariate time-series imputation, termed hybrid temporal modeling and selective spectral transport, which explicitly considers both temporal dependency learning and structural consistency preservation. Unlike conventional imputation methods that mainly rely on point-wise reconstruction in the time domain, the proposed framework combines long-short collaborative temporal modeling with selective spectral alignment in the frequency domain, thereby improving both recovery accuracy and temporal structural fidelity.

Experimental results on four benchmark datasets demonstrated that the proposed method achieves consistent and competitive improvements over representative baselines across different missing patterns and application domains. In addition, the downstream classification experiment on PhysioNet-2012 showed that the proposed method yields more informative reconstructed sequences for subsequent predictive modeling. The robustness evaluation under varying missing rates further confirmed the stability of the proposed framework, especially in challenging high-missingness scenarios.

These findings suggest that jointly modeling temporal dynamics and structural characteristics is a promising direction for time-series imputation. As future work, we will investigate more adaptive spectral alignment strategies, broader forms of structured missingness, and the extension of the proposed framework to other time-series analysis tasks.

References

[1] J. Wang, W. Du, W. Cao, K. Zhang, W. Wang, Y. Liang, and Q. Wen, "Deep learning for multivariate time series imputation: A survey," *CoRR*, vol. abs/2402.04059, 2024. [Online]. Available: https://dblp.org/rec/journals/corr/abs-2402-04059

[2] X. Miao, Y. Wu, L. Chen, Y. Gao, and J. Yin, "An experimental survey of missing data imputation algorithms," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 7, pp. 6630–6650, 2023. [Online]. Available: https://dblp.org/rec/journals/tkde/MiaoWCGY23

[3] Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, "Recurrent neural networks for multivariate time series with missing values," *CoRR*, vol. abs/1606.01865, 2016. [Online]. Available: https://dblp.org/rec/journals/corr/ChePCSL16

[4] W. Cao, D. Wang, J. Li, H. Zhou, L. Li, and Y. Li, "BRITS: Bidirectional recurrent imputation for time series," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2018, pp. 6775–6785. [Online]. Available: https://dblp.org/rec/conf/nips/CaoWLZLL18

[5] V. Fortuin, D. Baranchuk, G. Raetsch, and S. Mandt, "GP-VAE: Deep probabilistic time series imputation," in *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2020, pp. 1651–1661. [Online]. Available: https://dblp.org/rec/conf/aistats/FortuinBRM20

[6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *CoRR*, vol. abs/1706.03762, 2017. [Online]. Available: https://dblp.org/rec/journals/corr/VaswaniSPUJGKP17

[7] A. Gu, K. Goel, and C. Ré, "Efficiently modeling long sequences with structured state spaces," in *International Conference on Learning Representations (ICLR)*, 2022. [Online]. Available: https://dblp.org/rec/conf/iclr/GuGR22

[8] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," *CoRR*, vol. abs/2312.00752, 2023. [Online]. Available: https://dblp.org/rec/journals/corr/abs-2312-00752

[9] G. Peyré and M. Cuturi, "Computational optimal transport," *Foundations and Trends in Machine Learning*, vol. 11, no. 5-6, pp. 355–607, 2019. [Online]. Available: https://dblp.org/rec/journals/ftml/PeyreC19

[10] M. Cuturi, "Sinkhorn distances: Lightspeed computation of optimal transport," in *Advances in Neural Information Processing Systems (NIPS)*, 2013, pp. 2292–2300. [Online]. Available: https://dblp.org/rec/conf/nips/Cuturi13

[11] T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun, and R. Jin, "FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting," in *Proceedings of the 39th International Conference on Machine Learning (ICML)*, 2022, pp. 27 268–27 286. [Online]. Available: https://dblp.org/rec/conf/icml/ZhouMWW0022

[12] X. Yang, Y. Sun, X. Yuan, and X. Chen, "Frequency-aware generative models for multivariate time series imputation," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2024. [Online]. Available: https://dblp.org/rec/conf/nips/YangSYC24

[13] G. B. Moody, R. G. Mark, and A. L. Goldberger, "Physionet: Physiologic signals, time series and related open source software for basic, clinical, and applied research," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2011, pp. 8327–8330. [Online]. Available: https://dblp.org/rec/conf/embc/MoodyMG11

[14] S. Chen, "Beijing multi-site air-quality data," 2019. [Online]. Available: https://dblp.org/rec/data/10/Chen19f

[15] A. Trindade, "Electricityloaddiagrams20112014," 2015. [Online]. Available: https://dblp.org/rec/data/10/Trindade15

[16] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 12, 2021, pp. 11 106–11 115. [Online]. Available: https://dblp.org/rec/conf/aaai/ZhouZPZLXZ21

[17] Y. Tashiro, J. Song, Y. Song, and S. Ermon, "CSDI: Conditional score-based diffusion models for probabilistic time series imputation," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2021, pp. 24 804–24 816. [Online]. Available: https://dblp.org/rec/conf/nips/TashiroSSE21

[18] W. Du, D. Cote, and Y. Liu, "SAITS: Self-attention-based imputation for time series," *Expert Systems with Applications*, vol. 219, p. 119619, 2023. [Online]. Available: https://dblp.org/rec/journals/eswa/DuCL23

[19] Y. Luo, X. Cai, Y. Zhang, J. Xu, and X. Yuan, "Multivariate time series imputation with generative adversarial networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2018, pp. 1596–1607. [Online]. Available: https://dblp.org/rec/conf/nips/LuoCZXY18

[20] ——, "E2GAN: End-to-end generative adversarial network for multivariate time series imputation," in *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*, 2019, pp. 3094–3100. [Online]. Available: https://dblp.org/rec/conf/ijcai/Luo0CY19