Deep Learning for Multivariate Time Series Imputation: A Survey

Jun Wang[1,2,6†], **Wenjie Du**[1†], **Yiyuan Yang**[1,3], **Linglong Qian**[1,4], **Wei Cao**[1], **Keli Zhang**[5], **Wenjia Wang**[2,6], **Yuxuan Liang**[6], **Qingsong Wen**[7*]

[1]PyPOTS Research
[2]Hong Kong University of Science and Technology
[3]University of Oxford
[4]King's College London
[5]Huawei Noah's Ark Lab
[6]Hong Kong University of Science and Technology (Guangzhou)
[7]Squirrel Ai Learning

jwangfx@connect.ust.hk, wdu@time-series.ai, yiyuan.yang@cs.ox.ac.uk, linglong.qian@kcl.ac.uk, weicaomsra@gmail.com, zhangkeli1@huawei.com, wenjiawang@ust.hk, yuxliang@outlook.com, qingsongedu@gmail.com

Abstract

Missing values are ubiquitous in multivariate time series (MTS) data, posing significant challenges for accurate analysis and downstream applications. In recent years, deep learning-based methods have successfully handled missing data by leveraging complex temporal dependencies and learned data distributions. In this survey, we provide a comprehensive summary of deep learning approaches for multivariate time series imputation (MTSI) tasks. We propose a novel taxonomy that categorizes existing methods based on two key perspectives: imputation uncertainty and neural network architecture. Furthermore, we summarize existing MTSI toolkits with a particular emphasis on the PyPOTS Ecosystem, which provides an integrated and standardized foundation for MTSI research. Finally, we discuss key challenges and future research directions, which give insight for further MTSI research. This survey aims to serve as a valuable resource for researchers and practitioners in the field of time series analysis and missing data imputation tasks. A well-maintained MTSI paper and tool list is available at https://github.com/WenjieDu/Awesome_Imputation.

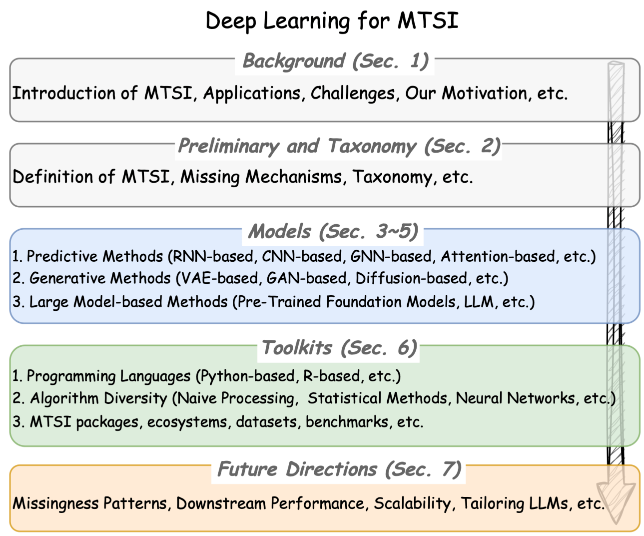

Figure 1: The framework of our survey.

1 Introduction

The data collection process of multivariate time series in various fields is often fraught with difficulties and uncertainty. In IoT systems, sensor failures and unstable environments lead to missing measurements [Li *et al.*, 2023]. Clinical studies face challenges from irregular sampling and privacy concerns [Ibrahim *et al.*, 2012; Esteban *et al.*, 2017; Qian *et al.*, 2025]. Financial and transportation systems encounter data gaps due to system downtime and communication issues [Bai and Ng, 2008; Gong *et al.*, 2021]. These missing values can significantly affect the accuracy and reliability of downstream analysis and decision-making. In real-world datasets like PhysioNet2012 [Silva *et al.*, 2012], missing rates can exceed 80%. Consequently, exploring how to reasonably and effectively impute missing components in multivariate time series is attractive and essential.

The earlier statistical imputation methods have historically been widely used for handling missing data. Those methods substitute the missing values with the statistics (e.g., zero value, mean value, and last observed value [Amiri and Jensen, 2016]) or simple statistical models, including ARIMA [Bartholomew, 1971], ARFIMA [Hamzaçebi, 2008], and SARIMA [Hamzaçebi, 2008]. Furthermore, machine learning techniques like regression, K-nearest neighbor, matrix factorization, etc., have gained prominence in the literature for addressing missing values in multivariate

*Corresponding author.

time series. Key implementations of these approaches include KNNI [Altman, 1992], TIDER [Liu *et al.*, 2022], MICE [Van Buuren and Groothuis-Oudshoorn, 2011], etc. While statistical and machine learning imputation methods are simple and efficient, they fall short in capturing the intricate temporal relationships and complex variation patterns inherent in time series data, resulting in limited performance.

More recently, deep learning-based imputation methods have demonstrated strong modelling capabilities for handling missing data. These approaches leverage advanced architectures such as Transformers, Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), diffusion models, Pre-trained Foundation Models (PFMs), and Large Language Models (LLMs) to capture the underlying structures and complex temporal dynamics of time series data. By learning the true data distribution from observed values, deep learning imputation methods can generate more reliable and contextually appropriate estimates for missing components. While several surveys exist on imputation techniques [Khayati *et al.*, 2020; Fang and Wang, 2020], they primarily focus on statistical and traditional machine learning approaches, offering limited discussion on deep learning-based methods. Given that multivariate time series imputation is a critical *data preprocessing* step for downstream time series analysis, a comprehensive and systematic survey on deep learning-driven imputation methods would provide valuable insights and contribute significantly to the research community.

In this survey paper, we endeavor to bridge the existing knowledge gap by providing a comprehensive summary of the latest developments in deep learning methods for multivariate time series imputation (MTSI). The framework of this survey is illustrated in Figure 1. In detail, we first present a succinct introduction to the topic, followed by the proposal of a novel taxonomy, categorizing approaches based on two perspectives: *imputation uncertainty* and *neural network architecture*. Imputation uncertainty quantifies the confidence in estimated values for missing data. To capture this, multiple stochastic samples are generated, and imputations are performed across these variations [Little and Rubin, 2019]. Accordingly, we categorize imputation methods into predictive ones, offering fixed estimates, and generative ones, which provide a distribution of possible values to account for imputation uncertainty. For neural network architecture, we explore a range of deep learning models tailored for MTSI, including RNN-based ones, GNN-based ones, CNN-based ones, attention-based ones, VAE-based ones, GAN-based ones, diffusion-based ones, PFM-based ones, and LLM-based ones.

To the best of our knowledge, this is the first comprehensive and systematic review of deep learning algorithms in the realm of MTSI, aiming to stimulate further research in this field. A corresponding resource that has been continuously updated can be found in our GitHub repository https://github.com/WenjieDu/Awesome_Imputation. In summary, the contributions of this survey include:

1. We introduce a novel taxonomy for deep multivariate time series imputation, and categorize them based on imputation uncertainty and neural network architecture.
2. We provide a comprehensive overview of existing MTSI toolkits. In particular, we highlight the PyPOTS Ecosystem, which integrates diverse imputation algorithms, standardized pipelines, and benchmarking resources, facilitating accessible and reproducible MTSI research.
3. We identify future directions, including missingness patterns, downstream task integration, and model scalability, offering insights to guide further advancements.

2 Preliminary and Taxonomy

2.1 Background of MTSI

Problem Definition A complete time-series dataset on $[0, T]$ typically can be denoted as $\mathcal{D} = \{\mathbf{X}_i, \mathbf{t}_i\}_{i=1}^N$. Hereby, $\mathbf{X}_i = \{x_{1:K,1:L}\} \in \mathcal{R}^{K \times L}$ and $\mathbf{t}_i = (t_1, \cdots, t_L) \in [0, T]^L$, where K is the number of features and L is the length of time series. In the missing data context, each complete time series can be split into an observed and a missing part, i.e., $\mathbf{X}_i = \{\mathbf{X}_i^o, \mathbf{X}_i^m\}$. For encoding the missingness, we also denote an observation matrix as $\mathbf{M}_i = \{m_{1:K,1:L}\}$, where $m_{k,l} = 0$ if $x_{k,l}$ is missing at timestamp t_l, otherwise $m_{k,l} = 1$. Furthermore, we can also calculate a time-lag matrix $\boldsymbol{\delta}_i = \{\delta_{1:K,1:L}\}$ by the following rule:

$$\delta_{k,l} = \begin{cases} 0, & \text{if } l = 1 \\ t_l - t_{l-1}, & \text{if } m_{k,l-1} = 1 \text{ and } l > 1 \\ \delta_{k,l-1} + t_l - t_{l-1}, & \text{if } m_{k,l-1} = 0 \text{ and } l > 1 \end{cases}$$

Hence, each incomplete time series is expressed as $\{\mathbf{X}_i^o, \mathbf{M}_i, \boldsymbol{\delta}_i\}$. The objective of MTSI is to construct an imputation model $\mathcal{M}_\theta$, parameterized by θ, to accurately estimate missingness in $\mathbf{X}^o$. The *imputed* matrix is defined as:

$$\hat{\mathbf{X}} = \mathbf{M} \odot \mathbf{X}^o + (1 - \mathbf{M}) \odot \bar{\mathbf{X}}, \tag{1}$$

where $\odot$ denotes element-wise multiplication, and $\bar{\mathbf{X}} = \mathcal{M}_\theta(\mathbf{X}^o)$ is the reconstructed matrix. The aim of $\mathcal{M}_\theta$ is twofold: (i) to make $\hat{\mathbf{X}}$ approximate the true *complete* data $\mathbf{X}$ as closely as possible, or (ii) to enhance the downstream task performance using $\hat{\mathbf{X}}$ compared to using the original $\mathbf{X}^o$.

Missing Mechanism The missing mechanisms, i.e., the cause of missing data, define the statistical relationship between observations and the probability of missing values [Nakagawa, 2015]. In real-world scenarios, missing mechanisms are inherently complex, and imputation model performance heavily depends on how well our assumptions align with the actual missing data patterns. According to Rubin's theory [Rubin, 1976], missing mechanisms fall into three categories: Missing Completely At Random (MCAR), Missing At Random (MAR), and Missing Not At Random (MNAR). MCAR implies that missingness is independent of both observed and missing data. Conversely, MAR indicates that missingness depends solely on observed data. MNAR suggests that missingness is related to the missing data itself and may also be influenced by observed data. These mechanisms can be formally defined as follows:

- MCAR: $p(\mathbf{M}|\mathbf{X}) = p(\mathbf{M})$,
- MAR: $p(\mathbf{M}|\mathbf{X}) = p(\mathbf{M}|\mathbf{X}^o)$,
- MNAR: $p(\mathbf{M}|\mathbf{X}) = p(\mathbf{M}|\mathbf{X}^o, \mathbf{X}^m)$.

MCAR and MAR are stronger assumptions compared to MNAR and are considered "ignorable" [Little and Rubin, 2019]. This means that the missing mechanism can be disregarded during imputation, focusing solely on learning the data distribution, i.e., $p(\mathbf{X}^o)$. In contrast, MNAR, often more reflective of real-life scenarios, is "non-ignorable", overlooking its missing mechanism can lead to biased parameter estimates. The objective here shifts to learning the joint distribution of the data and its missing mechanism, i.e., $p(\mathbf{X}^o, \mathbf{M})$.

2.2 Taxonomy of Deep Learning-based MTSI

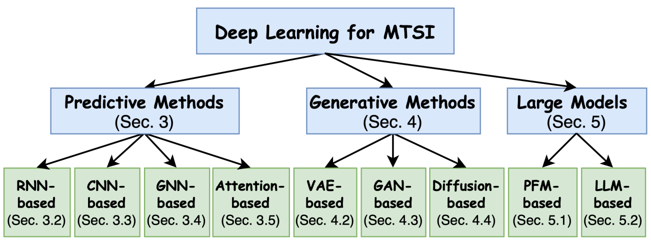

Figure 2: The taxonomy of deep learning methods for multivariate time series imputation from the view of imputation uncertainty and neural network architecture.

To summarize the existing deep multivariate time series imputation methods, we propose a taxonomy from the perspectives of imputation uncertainty and neural network architecture as illustrated in Figure 2, and provide a more detailed summary of these methods in Table 1. Regarding large model-based methods, we separate them as a category considering their application strategies are quite different from general neural network algorithms. For imputation uncertainty, we categorize imputation methods into predictive and generative types, based on their ability to yield varied imputations that reflect the inherent uncertainty in the imputation process. In the context of the neural network architecture, we examine prominent deep learning models specifically designed for multivariate time series imputation. The discussed models encompass RNN-based ones, CNN-based ones, GNN-based ones, attention-based ones, VAE-based ones, GAN-based ones, and diffusion-based ones. In the following sections, we will delve into and discuss the existing deep time series imputation methods from these two perspectives.

3 Predictive Methods

This section delves into predictive imputation methods, and our discussion primarily focuses on four types: RNN-based, CNN-based, GNN-based, and attention-based models.

3.1 Learning Objective

Predictive imputation methods consistently predict deterministic values for the same missing components, thereby not accounting for the uncertainty in the imputed values. Typically, these methods employ a reconstruction-based learning manner with the learning objective being,

$$\mathcal{L}_{det}(\theta) = \sum_{i=1}^{N} \ell_e(\mathbf{M}_i \odot \bar{\mathbf{X}}_i, \mathbf{M}_i \odot \mathbf{X}_i^o), \tag{2}$$

where ℓ_e is an absolute or squared error function.

3.2 RNN-based Models

As a natural way to model sequential data, Recurrent Neural Networks (RNNs) were developed early on the topic of advanced time-series analysis, and imputation is not an exception. GRU-D [Che *et al.*, 2018], a variant of GRU, is designed to process time series containing missing values. It is regulated by a temporal decay mechanism, which takes the time-lag matrix δ_i as input and models the temporal irregularity caused by missing values. Temporal belief memory [Kim and Chi, 2018], inspired by a biological neural model called the Hodgkin–Huxley model, is proposed to handle missing data by computing a belief of each feature's last observation with a bidirectional RNN and imputing a missing value based on its corresponding belief. M-RNN [Yoon *et al.*, 2019] is an RNN variant that works in a multi-directional way. This model interpolates within data streams with a bidirectional RNN model and imputes across data streams with a fully connected network. BRITS [Cao *et al.*, 2018] models incomplete time series with a bidirectional RNN. It takes missing values as variables of the RNN graph and fills in missing data with the hidden states from the RNN. In addition to imputation, BRITS is capable of working on the time series classification task simultaneously. Both M-RNN and BRITS adopt the temporal decay function from GRU-D to capture the informative missingness for performance improvement. Subsequent works, such as [Luo *et al.*, 2018; Luo *et al.*, 2019; Liu *et al.*, 2019; Miao *et al.*, 2021], combine RNNs with the GAN structure to output imputation with higher accuracy.

3.3 CNN-based Models

Convolutional Neural Networks (CNNs) represent a foundational deep learning architecture, extensively employed in sophisticated time series analysis. TimesNet [Wu *et al.*, 2023a] innovatively incorporates Fast Fourier Transform to restructure 1D time series into a 2D format, facilitating the utilization of CNNs for data processing. Also in GP-VAE [Fortuin *et al.*, 2020], CNNs play the role of the backbone in both the encoder and decoder. Furthermore, CNNs serve as pivotal feature extractors within attention-based models like DeepMVI [Bansal *et al.*, 2021], as well as in diffusion-based models such as CSDI [Tashiro *et al.*, 2021], by mapping input data into an embedding space for subsequent processing.

3.4 GNN-based Models

GNN-based models treat time series as graph sequences, reconstructing missing values via learned node representations. [Cini *et al.*, 2022] introduces GRIN, the first graph-based recurrent architecture for MTSI, leveraging a bidirectional graph recurrent neural network to capture temporal dynamics and spatial similarities, significantly improving imputation accuracy. SPIN [Marisca *et al.*, 2022] further integrates a sparse spatiotemporal attention mechanism into the GNN framework, mitigating GRIN's error propagation and enhancing robustness against data sparsity.

	Imputation Uncertainty	Neural Network Architecture	Missing Mechanism		
GRU-D [Che *et al.*, 2018]	Scientific Reports	predictive	✖	RNN	MCAR
M-RNN [Yoon *et al.*, 2019]	TBME	predictive	✖	RNN	MCAR
BRITS [Cao *et al.*, 2018]	NeurIPS	predictive	✖	RNN	MCAR
TimesNet [Wu *et al.*, 2023a]	ICLR	predictive	✖	CNN	MCAR
GRIN [Cini *et al.*, 2022]	ICLR	predictive	✖	GNN	MCAR / MAR
SPIN [Marisca *et al.*, 2022]	NeurIPS	predictive	✖	GNN, Attention	MCAR / MAR
Transformer [Vaswani *et al.*, 2017]	NeurIPS	predictive	✖	Attention	MCAR
SAITS [Du *et al.*, 2023]	ESWA	predictive	✖	Attention	MCAR
DeepMVI [Bansal *et al.*, 2021]	VLDB	predictive	✖	Attention, CNN	MCAR
ImputeFormer [Nie *et al.*, 2024]	KDD	predictive	✖	Attention	MCAR
Casper [Jing *et al.*, 2024]	CIKM	predictive	✖	GNN, Attention	MCAR
HSPGNN [Liang *et al.*, 2024a]	CIKM	predictive	✖	GNN	MCAR
GP-VAE [Fortuin *et al.*, 2020]	AISTATS	generative	✔⃝	VAE, CNN	MCAR / MAR
V-RIN [Mulyadi *et al.*, 2021]	Trans. Cybern.	generative	✔	VAE, RNN	MCAR / MAR
supnotMIWAE [Kim *et al.*, 2023]	ICML	generative	✔⃝	VAE	MNAR
GRUI-GAN [Luo *et al.*, 2018]	NeurIPS	generative	✔⃝	GAN, RNN	MCAR
E^{2}GAN [Luo *et al.*, 2019]	IJCAI	generative	✔⃝	GAN, RNN	MCAR
NAOMI [Liu *et al.*, 2019]	NeurIPS	generative	✔⃝	GAN, RNN	MCAR
SSGAN [Miao *et al.*, 2021]	AAAI	generative	✔⃝	GAN, RNN	MCAR
CSDI [Tashiro *et al.*, 2021]	NeurIPS	generative	✔⃝	Diffusion, Attention, CNN	MCAR
SSSD [Alcaraz and Strodthoff, 2023]	TMLR	generative	✔⃝	Diffusion, Attention	MCAR
CSBI [Chen *et al.*, 2023]	ICML	generative	✔⃝	Diffusion, Attention	MCAR
MIDM [Wang *et al.*, 2023]	KDD	generative	✔⃝	Diffusion, Attention	MCAR
PriSTI [Liu *et al.*, 2023]	ICDE	generative	✔⃝	Diffusion, Attention, GNN, CNN	MCAR
SPD [Biloš *et al.*, 2023]	ICML	generative	✔⃝	Diffusion, Attention	MCAR
SADI [Dai *et al.*, 2024]	AISTATS	generative	✔	Diffusion, Attention	MCAR / MAR / MNAR
FGTI [Yang *et al.*, 2024a]	NeurIPS	generative	✔	Diffusion, Attention	MCAR / MAR / MNAR
MTSCI [Zhou *et al.*, 2024]	CIKM	generative	✔	Diffusion, Attention	MCAR
MOMENT [Goswami *et al.*, 2024]	ICML	large model	✔	Foundation model	MCAR
Timer [Liu *et al.*, 2024]	ICML	large model	✔	Foundation model	MCAR
Timemixer++ [Wang *et al.*, 2024]	ICLR	large model	✔	Foundation model	MCAR
GPT4TS [Zhou *et al.*, 2023]	NeurIPS	large model	✔	Large language model	MCAR
LLM-TS Integrator [Chen *et al.*, 2024]	NeurIPS workshop	large model	✔	Large language model	MCAR

Table 1: Summary of deep learning methods for multivariate time series imputation. ✔ and ✔⃝ indicate methods capable of accounting for imputation uncertainty, whereas ✖ denotes methods that do not. Furthermore, ✔ denotes that the methods also define the fidelity score to explicitly measure the imputation uncertainty.

3.5 Attention-based Models

Since Transformer is proposed in [Vaswani *et al.*, 2017], the self-attention mechanism has been widely used to model sequence data including time series [Wen *et al.*, 2023]. DeepMVI [Bansal *et al.*, 2021] integrates transformers with convolutional techniques, tailoring key-query designs to effectively address missing value imputation. For each time series, DeepMVI harnesses attention mechanisms to concurrently distill long-term seasonal, granular local, and cross-dimensional embeddings, which are concatenated to predict the final output. SAITS [Du *et al.*, 2023] employs a self-supervised training scheme to deal with missing data, which integrates dual joint learning tasks: a masked imputation task and an observed reconstruction task. This method, featuring two diagonal-masked self-attention blocks and a weighted-combination block, leverages attention weights and missingness indicators to enhance imputation precision. Besides, ImputeFormer [Nie *et al.*, 2024] introduces a novel Transformer-based framework that leverages self-attention and temporal context modeling to accurately recover missing values. In addition to the above models, the attention mechanism is also widely adapted to build the denoising network in diffusion models like CSDI [Tashiro *et al.*, 2021], MIDM [Wang *et al.*, 2023], PriSTI [Liu *et al.*, 2023], Diffusion-TS [Yuan and Qiao, 2024] and in GNN-based models like Casper [Jing *et al.*, 2024] and HSPGNN [Liang *et al.*, 2024a].

3.6 Pros and Cons

This subsection synthesizes the strengths and challenges of the predictive imputation methods discussed. RNN-based models, while adept at capturing sequential information, are inherently limited by their sequential processing nature and memory constraints, which may lead to scalability issues with long sequences [Khayati *et al.*, 2020]. Although CNNs have decades of development and are useful feature extractors to capture neighborhood information and local connectivity, their kernel size and working mechanism intrinsically limit their performance on time-series data as the backbone. Due to the attention mechanism, attention-based models generally outperform RNN-based and CNN-based methods in imputation tasks due to their superior ability to handle long-range dependencies and parallel processing capabilities. GNN-based methods provide a deeper understanding of spatio-temporal dynamics, yet they often come with increased computational complexity, posing challenges for large-scale or high-dimensional data.

4 Generative Methods

In this section, we examine generative imputation methods, including three primary types: VAE-based, GAN-based, and diffusion-based models.

4.1 Learning Objective

Generative methods are essentially built upon generative models like VAEs, GANs, and diffusion models. They are characterized by their ability to generate varied outputs for missing observations, enabling the quantification of imputation uncertainty. Typically, these methods learn probability distributions from the observed data and subsequently generate slightly different values aligned with these learned distributions for the missing observation. The primary learning objective of generative methods is thus defined as,

$$\mathcal{L}_{pro}(\theta) = \sum_{i=1}^{N} \log p_\theta(\mathbf{X}_i^o). \tag{3}$$

where θ is the model parameters of the imputation model $\mathcal{M}$.

4.2 VAE-based Models

VAEs employ an encoder-decoder structure to approximate the true data distribution by maximizing the Evidence Lower Bound (ELBO) on the marginal likelihood. This ELBO enforces a Gaussian-distributed latent space from which the decoder reconstructs diverse data points.

The authors in [Fortuin *et al.*, 2020] propose the first VAE-based imputation method GP-VAE, where they utilized a Gaussian process prior in the latent space to capture temporal dynamics. Moreover, the ELBO in GP-VAE is only evaluated on the observed features of the data. Authors in [Mulyadi *et al.*, 2021] design V-RIN to mitigate the risk of biased estimates in missing value imputation. V-RIN captures uncertainty by accommodating a Gaussian distribution over the model output, specifically interpreting the variance of the reconstructed data from a VAE model as an uncertainty measure. It then models temporal dynamics and seamlessly integrates this uncertainty into the imputed data through an uncertainty-aware GRU. More recently, authors in [Kim *et al.*, 2023] propose supnotMIWAE and introduce an extra classifier, where they extend the ELBO in GP-VAE to model the joint distribution of the observed data, its mask matrix, and its label. In this way, their ELBO effectively models the imputation uncertainty, and the additional classifier encourages the VAE model to produce missing values that are more advantageous for the downstream classification task.

4.3 GAN-based Models

GANs facilitate adversarial training through a minimax game between two components: a generator aiming to mimic the real data distribution, and a discriminator tasked with distinguishing between the generated and real data. This dynamic fosters a progressive refinement of synthetic data that increasingly resembles real samples.

In [Luo *et al.*, 2018], authors propose a two-stage GAN imputation method (GRUI-GAN), which is the first GAN-based method for imputing time-series data. GRUI-GAN first learns the distribution of the observed multivariate time-series data by a standard adversarial training manner and then optimizes the input noise of the generator to further maximize the similarity of the generated and observed multivariate time series. However, the second stage in GRUI-GAN needs a lot of time to find the best matched input vector, and this vector is not always the best especially when the initial value of the "noise" is not properly set. Then, an end-to-end GAN imputation model E^2GAN [Luo *et al.*, 2019] is further proposed, where the generator takes a denoising autoencoder module to avoid the "noise" optimization stage in GRUI-GAN. Meanwhile, authors in [Liu *et al.*, 2019] propose a non-autoregressive multi-resolution GAN model (NAOMI), where the generator is assembled by a forward-backward encoder and a multiresolution decoder. The imputed data are recursively generated by the multiresolution decoder in a non-autoregressive manner, which mitigates error accumulation in scenarios involving high-missing and long sequence time series. On the other hand, in [Miao *et al.*, 2021], authors propose USGAN, which generates high-quality imputed data by integrating a discriminator with a temporal reminder matrix. This matrix introduces added complexity to the training of the discriminator and subsequently leads to improvements in the generator's performance. Furthermore, they extend USGAN to a semi-supervised model SSGAN, by introducing an extra classifier. In this way, SSGAN leverages label information, allowing the generator to estimate missing values while conditioning on observed components and labels simultaneously.

4.4 Diffusion-based Models

As an emerging and potent category of generative models, diffusion models are adept at capturing complex data distributions by progressively adding and then reversing noise through a Markov chain of diffusion steps. Distinct from VAE, these models utilize a fixed training procedure and operate with high-dimensional latent variables that retain the dimensionality of the input data [Yang *et al.*, 2024b].

CSDI, introduced in [Tashiro *et al.*, 2021], stands out as the pioneering diffusion model specifically designed for MTSI. Different from conventional diffusion models, CSDI adopts a conditioned training approach, where a subset of observed data is utilized as conditional information to facilitate the generation of the remaining segment of observed data. However, the denoising network in CSDI relies on two transformers, exhibiting quadratic complexity concerning the number of variables and the time series length. This design limitation raises concerns about memory constraints, particularly when modeling extensive multivariate time series. In response to this challenge, a subsequent work by [Alcaraz and Strodthoff, 2023] introduces SSSD, which addresses the quadratic complexity issue by replacing transformers with structured state space models [Gu *et al.*, 2022]. This modification proves advantageous, especially when handling lengthy multivariate time series, as it mitigates the risk of memory overflow. Another approach CSBI, introduced in [Chen *et al.*, 2023], improves the efficiency by modeling the diffusion process as a Schrodinger bridge problem, which could be transformed into computation-friendly stochastic differential equations. Also, SADI [Dai *et al.*, 2024] is a similarity-aware diffusion model that leverages a self-attention mechanism to capture inter-patient similarities for effective imputation of missing values. MTSCI [Zhou *et al.*, 2024] leverages cross-channel correlations and multi-scale temporal dynamics features to effectively recover missing values.

Moreover, the efficacy of diffusion models is notably influenced by the construction and utilization of conditional information. MIDM [Wang *et al.*, 2023] proposes to sample noise from a distribution conditional on observed data's representations in the denoising process, In this way, it can explicitly preserve the intrinsic correlations between observed and missing data. PriSTI [Liu *et al.*, 2023] introduces the spatiotemporal dependencies as conditional information, i.e., provides the denoising network with spatiotemporal attention weights calculated by the conditional feature for spatiotemporal imputation. Besides, FGTI [Yang *et al.*, 2024a] is a frequency-guided framework that leverages spectral analysis to effectively capture both global periodic patterns and local temporal dynamics, thereby enhancing missing data recovery.

Contrasting with the above diffusion-based methods that treat time series as discrete time steps, SPD [Biloš *et al.*, 2023] views time series as discrete realizations of an underlying continuous function and generates data for imputation using stochastic process diffusion. In this way, SPD posits the continuous noise process as an inductive bias for the irregular time series, so as to better capture the true generative process, especially with the inherent stochasticity of the data.

4.5 Pros and Cons

This subsection delineates the advantages and limitations of the aforementioned generative imputation models. VAE-based models are adept at modeling probabilities explicitly and offering a theoretical foundation for understanding data distributions. However, they are often constrained by their generative capacity, which can limit their performance in capturing complex data variability. GAN-based models, on the other hand, excel in data generation, providing high-quality imputations with impressive fidelity to the original data distributions. Yet, they are notoriously challenging to train due to issues like vanishing gradients [Wu *et al.*, 2023b], which can hamper model stability and convergence. Diffusion-based models emerge as powerful generative tools with a strong capacity for capturing intricate data patterns. Nevertheless, their computational complexity is considerable, and they also suffer from issues related to boundary coherence between missing and observed parts [Lugmayr *et al.*, 2022].

5 Large Model-based Methods

Large models aim to tackle three critical challenges in MTSI tasks: complex temporal dependencies across multiple scales, diverse missingness patterns, and the need for robust generalization with limited domain data. In this section, we examine how pre-trained foundation models (PFMs) and large language models (LLMs) approach these challenges through distinct but complementary paths.

5.1 Pre-Trained Foundation Models

PFMs leverage large-scale pretraining on diverse datasets to enhance generalization and adaptability [Liang *et al.*, 2024b]. For example, by pretraining on a vast collection of multivariate time series, Timer [Liu *et al.*, 2024] learns a rich set of temporal representations, enabling it to generalize across different domains. It employs a self-supervised contrastive learning framework that refines its ability to reconstruct missing values while maintaining consistency in long-range dependencies. Building on the idea of large-scale pretraining, MOMENT [Goswami *et al.*, 2024] introduces the Time Series Pile and facilitates pretraining of high-capacity Transformer models using masked time series prediction tasks. On the other hand, NuwaTS [Cheng *et al.*,] repurposes pre-trained language models for time series imputation, utilizing specialized embeddings and contrastive learning to handle various missing data patterns across domains. Meanwhile, Timemixer++ [Wang *et al.*, 2024] explores an alternative to Transformer-based architectures by adopting token-mixing techniques. By mixing temporal and feature-wise representations, Timemixer++ effectively captures both short- and long-term dependencies while reducing computational overhead. These PFMs represent a significant step forward in MTSI, offering scalable, generalizable, and high-performance solutions that adapt to various missing patterns across domains.

5.2 Large Language Models

LLMs have shown good capabilities in sequential modeling due to their autoregressive nature, which enables them to capture sequence dependencies through next-token prediction. Their extensive model parameters equip them with a strong learning capacity, making them suitable for handling MTSI tasks [Jin *et al.*, 2023]. For example, GPT4TS [Zhou *et al.*, 2023] optimizes GPT-2 for MTSI by introducing a key architectural adjustment—the freezing of the attention module. This design allows the model to focus on fine-tuning positional embeddings and layer normalization layers using a small number of samples. Such targeted fine-tuning significantly improves the model's ability to reconstruct missing values with minimal data availability. Furthermore, LLM-TS Integrator [Chen *et al.*, 2024] enhances LLM-based MTSI tasks by integrating statistical and deep learning methods within a hybrid framework. It introduces an adaptive retrieval mechanism to select relevant historical patterns and a self-correction module for iterative refinement. By leveraging retrieval-augmented generation, the model effectively handles irregular and sparse time-series data.

These large model approaches represent a fundamental shift in time series imputation: from treating missing values as isolated points to understanding them within a rich temporal and cross-variable context. While showing promising results, particularly in scenarios with complex missingness patterns, their adoption requires careful consideration of computational resources and domain-specific constraints.

6 Time Series Imputation Toolkits

On the MTSI task, there are existing libraries providing naive processing ways, statistical methods, machine learning imputation algorithms, and deep learning imputation neural networks for convenient usage.

`imputeTS` [Moritz and Bartz-Beielstein, 2017], a library in R provides several naive approaches (e.g., mean values, last observation carried forward, etc.) and commonly-used imputation algorithms (e.g., linear interpolation, Kalman smoothing, and weighted moving average) but only for univariate time series. Another well-known R package, **`mice`** [Van Buuren and Groothuis-Oudshoorn, 2011], implements the method called multivariate imputation by chained equations to tackle missingness in data. Although it is not for time series specifically, it is widely used in practice for multivariate time-series imputation, especially in the field of statistics. **`Impyute`** (https://github.com/eltonlaw/impyute) and **`Autoimpute`** (https://github.com/kearnz/autoimpute) both offer naive imputation methods for cross-sectional data and time-series data. **`Impyute`** is only with simple approaches like the moving average window, and

`Autoimpute` integrates parametric methods, for example, polynomial interpolation and spline interpolation. More recently, **`GluonTS`** [Alexandrov *et al.*, 2020], a generative machine-learning package for time series, provides some naive ways, such as dummy value imputation and casual mean value imputation, to handle missing values. In addition to simple and non-parametric methods, **`Sktime`** [Löning *et al.*, 2019] implements one more option that allows users to leverage integrated machine learning imputation algorithms to predict missing values in the given data, though this works in a univariate way. **`ImputeBench`** [Khayati *et al.*, 2020] offers a collection of machine learning and deep learning-based imputation but lacks uniform programming languages.

When it comes to deep-learning imputation, **`PyPOTS`** [Du, 2023] is a toolbox focusing on modeling partially-observed time series end-to-end, which contains dozens of neural networks for tasks on incomplete time series, including 43 imputation models so far. Leveraging the kits from PyPOTS Ecosystem (https://pypots.com/ecosystem) that we developed, **TSI-Bench Suite** [Du *et al.*, 2024] provides a set of standard pipelines processing 172 public time-series datasets for benchmarking imputation algorithms. The comprehensive benchmark results from 34,804 experiments, including 28 algorithms, 8 typical datasets from different domains, and diverse missingness patterns are presented in [Du *et al.*, 2024].

7 Future Direction

Missingness Patterns Existing imputation algorithms predominantly operate under the MCAR or MAR assumptions. However, real-world missing data mechanisms are often more complex, with MNAR being prevalent across various domains. The non-ignorable nature of MNAR indicates a fundamental distributional shift between observed and true data. For example, in airflow signal analysis, the absence of high-value observations leads to saturated peaks, visibly skewing the observed data distribution compared to the true underlying one. This scenario illustrates how imputation methods may incur inductive bias in model parameter estimation and underperform in the presence of MNAR. Addressing missing data in MNAR contexts, distinct from MCAR and MAR, calls for innovative methodologies to achieve better performance.

Downstream Performance The primary objective of imputing missing values lies in enhancing downstream data analytics, particularly in scenarios with incomplete information. The prevalent approach is the "*impute and predict*" **two-stage paradigm**, where missing value imputation is a part of data preprocessing and followed by task-specific downstream models (e.g. a classifier), either in tandem or sequentially. An alternative method is the "*encode and predict*" **end-to-end paradigm**, encoding the incomplete data into a proper representation for multitask learning, including imputation and other tasks (e.g. classification and forecasting, etc.). Despite the optimal paradigm for partially-observed time series still remains an open area for future investigation, the latter end-to-end way turns out to be more promising especially when information embedded in the missing patterns is helpful to the downstream tasks [Miyaguchi *et al.*, 2021].

Scalability While deep learning imputation algorithms have shown impressive performance, their computational cost often exceeds that of statistical and machine learning-based counterparts. In the era of burgeoning digital data, spurred by advancements in communication and IoT devices, we are witnessing an exponential increase in data generation. This surge, accompanied by the prevalence of incomplete datasets, poses significant challenges in training deep models effectively [Wu *et al.*, 2023b]. Specifically, the high computational demands of existing deep imputation algorithms render them less feasible for large-scale datasets. Consequently, there is a growing need for scalable deep imputation solutions, leveraging parallel and distributed computing techniques, to effectively address the challenges of large-scale missing data.

Large Language Models for MTSI While large models have shown promising results in time series imputation, several critical research directions remain unexplored. First, beyond current architectural innovations, the explicit incorporation of domain-specific temporal constraints and prior knowledge about missingness mechanisms into the pretraining process offers significant potential. Second, while the latest models have advanced temporal modeling, there remains room for fundamental innovations in processing irregular temporal patterns, particularly through more efficient and interpretable architectures. Third, the potential of multimodal learning deserves further investigation, where large models' ability to process different data modalities could incorporate auxiliary information (such as textual descriptions or metadata) to achieve more accurate and contextually appropriate imputations. These directions, coupled with robust evaluation frameworks assessing uncertainty quantification and temporal consistency, could significantly advance time series imputation in critical applications.

8 Conclusion

This survey presents a systematic review of deep learning-based methods for multivariate time series imputation with a novel taxonomy to categorize predictive and generative methods and also discusses the large model for MTSI tasks. We provide a comprehensive architecture overview, highlighting their strengths, limitations, and applications. To advance this field, we identify key challenges, including handling MNAR missingness, integrating imputation with downstream tasks, and improving scalability. Future research should explore large-scale pre-trained models and multimodal learning to enhance robustness and real-world applicability.

Contribution Statement

Jun Wang and Wenjie Du contribute equally and are co-first authors.

References

[Alcaraz and Strodthoff, 2023] Juan Lopez Alcaraz and Nils Strodthoff. Diffusion-based time series imputation and forecasting with structured state space models. *Transactions on Machine Learning Research*, 2023.

[Alexandrov *et al.*, 2020] Alexander Alexandrov, Konstantinos Benidis, Michael Bohlke-Schneider, Valentin Flunkert, Jan Gasthaus, et al. GluonTS: Probabilistic and Neural Time Series Modeling in Python. *Journal of Machine Learning Research*, 21(116):1–6, 2020.

[Altman, 1992] Naomi S Altman. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185, 1992.

[Amiri and Jensen, 2016] Mehran Amiri and Richard Jensen. Missing data imputation using fuzzy-rough methods. *Neurocomputing*, 205(1):152–164, 2016.

[Bai and Ng, 2008] Jushan Bai and Serena Ng. Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2):304–317, 2008.

[Bansal *et al.*, 2021] Parikshit Bansal, Prathamesh Deshpande, and Sunita Sarawagi. Missing value imputation on multidimensional time series. In *VLDB*, 2021.

[Bartholomew, 1971] David J Bartholomew. Time series analysis forecasting and control. *Journal of the Operational Research Society*, 22(2):199–201, 1971.

[Biloš *et al.*, 2023] Marin Biloš, Kashif Rasul, Anderson Schneider, Yuriy Nevmyvaka, and Stephan Günnemann. Modeling temporal data as continuous functions with stochastic process diffusion. In *ICML*, 2023.

[Cao *et al.*, 2018] Wei Cao, Dong Wang, Jian Li, Hao Zhou, Lei Li, and Yitan Li. Brits: Bidirectional recurrent imputation for time series. *NeurIPS*, 2018.

[Che *et al.*, 2018] Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, David Sontag, and Yan Liu. Recurrent neural networks for multivariate time series with missing values. *Scientific Reports*, 8(1), Apr 2018.

[Chen *et al.*, 2023] Yu Chen, Wei Deng, Shikai Fang, Fengpei Li, Nicole Tianjiao Yang, Yikai Zhang, et al. Provably convergent schrödinger bridge with applications to probabilistic time series imputation. In *ICML*, 2023.

[Chen *et al.*, 2024] Can Chen, Gabriel L Oliveira, Hossein Sharifi-Noghabi, and Tristan Sylvain. Enhance time series modeling by integrating llm. In *NeurIPS Workshop on Time Series in the Age of Large Models*, 2024.

[Cheng *et al.*,] Jinguo Cheng, Chunwei Yang, Wanlin Cai, Yuxuan Liang, Qingsong Wen, and Yuankai Wu. Nuwats: a foundation model mending every incomplete time series.

[Cini *et al.*, 2022] Andrea Cini, Ivan Marisca, and Cesare Alippi. Filling the g_ap_s: Multivariate time series imputation by graph neural networks. In *ICLR*, 2022.

[Dai *et al.*, 2024] Zongyu Dai, Emily Getzen, and Qi Long. Sadi: Similarity-aware diffusion model-based imputation for incomplete temporal ehr data. In *International Conference on Artificial Intelligence and Statistics*, pages 4195–4203. PMLR, 2024.

[Du *et al.*, 2023] Wenjie Du, David Cote, and Yan Liu. SAITS: Self-Attention-based Imputation for Time Series. *Expert Systems with Applications*, 219:119619, 2023.

[Du *et al.*, 2024] Wenjie Du, Jun Wang, Linglong Qian, Yiyuan Yang, Zina Ibrahim, Fanxing Liu, Zepu Wang, Haoxin Liu, Zhiyuan Zhao, Yingjie Zhou, et al. Tsi-bench: Benchmarking time series imputation. *arXiv preprint arXiv:2406.12747*, 2024.

[Du, 2023] Wenjie Du. PyPOTS: a Python toolbox for data mining on Partially-Observed Time Series. In *SIGKDD workshop on Mining and Learning from Time Series*, 2023.

[Esteban *et al.*, 2017] Cristóbal Esteban, Stephanie L Hyland, and Gunnar Rätsch. Real-valued (medical) time series generation with recurrent conditional gans. *ArXiv Preprint ArXiv:1706.02633*, 2017.

[Fang and Wang, 2020] Chenguang Fang and Chen Wang. Time series data imputation: A survey on deep learning approaches. *arXiv preprint arXiv:2011.11347*, 2020.

[Fortuin *et al.*, 2020] Vincent Fortuin, Dmitry Baranchuk, Gunnar Raetsch, and Stephan Mandt. GP-VAE: Deep probabilistic time series imputation. In *AISTATS*, 2020.

[Gong *et al.*, 2021] Yongshun Gong, Zhibin Li, Jian Zhang, Wei Liu, Yilong Yin, and Yu Zheng. Missing value imputation for multi-view urban statistical data via spatial correlation learning. *IEEE Transactions on Knowledge and Data Engineering*, 2021.

[Goswami *et al.*, 2024] Mononito Goswami, Konrad Szafer, Arjun Choudhry, Yifu Cai, Shuo Li, and Artur Dubrawski. Moment: A family of open time-series foundation models. In *ICML*, 2024.

[Gu *et al.*, 2022] Albert Gu, Karan Goel, and Christopher Re. Efficiently modeling long sequences with structured state spaces. In *ICLR*, 2022.

[Hamzaçebi, 2008] Coşkun Hamzaçebi. Improving artificial neural networks' performance in seasonal time series forecasting. *Information Sciences*, 178(23):4550–4559, 2008.

[Ibrahim *et al.*, 2012] Joseph G Ibrahim, Haitao Chu, and Ming-Hui Chen. Missing data in clinical studies: issues and methods. *Journal of clinical oncology*, 2012.

[Jin *et al.*, 2023] Ming Jin, Qingsong Wen, Yuxuan Liang, Chaoli Zhang, Siqiao Xue, Xue Wang, James Zhang, Yi Wang, Haifeng Chen, Xiaoli Li, et al. Large models for time series and spatio-temporal data: A survey and outlook. *arXiv preprint arXiv:2310.10196*, 2023.

[Jing *et al.*, 2024] Baoyu Jing, Dawei Zhou, Kan Ren, and Carl Yang. Causality-aware spatiotemporal graph neural networks for spatiotemporal time series imputation. In *CIKM*, 2024.

[Khayati *et al.*, 2020] Mourad Khayati, Alberto Lerner, Zakhar Tymchenko, and Philippe Cudré-Mauroux. Mind the gap: an experimental evaluation of imputation of missing values techniques in time series. In *VLDB*, 2020.

[Kim and Chi, 2018] Yeo Jin Kim and Min Chi. Temporal Belief Memory: Imputing missing data during rnn training. In *IJCAI*, 2018.

[Kim *et al.*, 2023] Seunghyun Kim, Hyunsu Kim, Eunggu Yun, Hwangrae Lee, Jaehun Lee, and Juho Lee. Probabilistic imputation for time-series classification with missing data. In *ICML*, 2023.

[Li *et al.*, 2023] Xiao Li, Huan Li, Harry Kai-Ho Chan, Hua Lu, and Christian S Jensen. Data imputation for sparse radio maps in indoor positioning. In *ICDE*, 2023.

[Liang *et al.*, 2024a] Guojun Liang, Prayag Tiwari, Sławomir Nowaczyk, and Stefan Byttner. Higher-order spatio-temporal physics-incorporated graph neural network for multivariate time series imputation. In *CIKM*, 2024.

[Liang *et al.*, 2024b] Yuxuan Liang, Haomin Wen, Yuqi Nie, Yushan Jiang, Ming Jin, Dongjin Song, Shirui Pan, and Qingsong Wen. Foundation models for time series analysis: A tutorial and survey. In *KDD*, 2024.

[Little and Rubin, 2019] Roderick JA Little and Donald B Rubin. *Statistical analysis with missing data*, volume 793. John Wiley & Sons, 2019.

[Liu *et al.*, 2019] Yukai Liu, Rose Yu, Stephan Zheng, Eric Zhan, and Yisong Yue. Naomi: Non-autoregressive multiresolution sequence imputation. In *NeurIPS*, 2019.

[Liu *et al.*, 2022] Shuai Liu, Xiucheng Li, Gao Cong, Yile Chen, and Yue Jiang. Multivariate time-series imputation with disentangled temporal representations. 2022.

[Liu *et al.*, 2023] Mingzhe Liu, Han Huang, Hao Feng, Leilei Sun, Bowen Du, and Yanjie Fu. Pristi: A conditional diffusion framework for spatiotemporal imputation. *arXiv preprint arXiv:2302.09746*, 2023.

[Liu *et al.*, 2024] Yong Liu, Haoran Zhang, Chenyu Li, Xiangdong Huang, Jianmin Wang, and Mingsheng Long. Timer: Generative pre-trained transformers are large time series models. In *Forty-first International Conference on Machine Learning*, 2024.

[Löning *et al.*, 2019] Markus Löning, Anthony Bagnall, Sajaysurya Ganesh, Viktor Kazakov, Jason Lines, et al. sktime: A unified interface for machine learning with time series. *arXiv preprint arXiv:1909.07872*, 2019.

[Lugmayr *et al.*, 2022] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *CVPR*, 2022.

[Luo *et al.*, 2018] Yonghong Luo, Xiangrui Cai, Ying ZHANG, Jun Xu, and Yuan xiaojie. Multivariate time series imputation with generative adversarial networks. In *NeurIPS*, 2018.

[Luo *et al.*, 2019] Yonghong Luo, Ying Zhang, Xiangrui Cai, and Xiaojie Yuan. E^2GAN: End-to-end generative adversarial network for multivariate time series imputation. In *IJCAI*, 2019.

[Marisca *et al.*, 2022] Ivan Marisca, Andrea Cini, and Cesare Alippi. Learning to reconstruct missing data from spatiotemporal graphs with sparse observations. *NeurIPS*, 2022.

[Miao *et al.*, 2021] Xiaoye Miao, Yangyang Wu, Jun Wang, Yunjun Gao, Xudong Mao, and Jianwei Yin. Generative semi-supervised learning for multivariate time series imputation. In *AAAI*, 2021.

[Miyaguchi *et al.*, 2021] Kohei Miyaguchi, Takayuki Katsuki, Akira Koseki, and Toshiya Iwamori. Variational inference for discriminative learning with generative modeling of feature incompletion. In *ICLR*, 2021.

[Moritz and Bartz-Beielstein, 2017] Steffen Moritz and Thomas Bartz-Beielstein. imputeTS: Time Series Missing Value Imputation in R. *The R Journal*, 2017.

[Mulyadi *et al.*, 2021] Ahmad Wisnu Mulyadi, Eunji Jun, and Heung-Il Suk. Uncertainty-aware variational-recurrent imputation network for clinical time series. *IEEE Transactions on Cybernetics*, 52(9):9684–9694, 2021.

[Nakagawa, 2015] Shinichi Nakagawa. Missing data: mechanisms, methods and messages. pages 81–105. Oxford University Press Oxford, UK, 2015.

[Nie *et al.*, 2024] Tong Nie, Guoyang Qin, Wei Ma, Yuewen Mei, and Jian Sun. Imputeformer: Low rankness-induced transformers for generalizable spatiotemporal imputation. In *KDD*, 2024.

[Qian *et al.*, 2025] Linglong Qian, Hugh Logan Ellis, Tao Wang, Jun Wang, Robin Mitra, Richard Dobson, and Zina Ibrahim. How deep is your guess? a fresh perspective on deep learning for medical time-series imputation. *IEEE Journal of Biomedical and Health Informatics*, 2025.

[Rubin, 1976] Donald B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 12 1976.

[Silva *et al.*, 2012] Ikaro Silva, George Moody, Daniel J Scott, Leo A Celi, and Roger G Mark. Predicting in-hospital mortality of icu patients: The physionet/computing in cardiology challenge 2012. *Computing in cardiology*, 39:245, 2012.

[Tashiro *et al.*, 2021] Yusuke Tashiro, Jiaming Song, Yang Song, and Stefano Ermon. CSDI: Conditional score-based diffusion models for probabilistic time series imputation. In *NeurIPS*, 2021.

[Van Buuren and Groothuis-Oudshoorn, 2011] Stef Van Buuren and Karin Groothuis-Oudshoorn. mice: Multivariate imputation by chained equations in r. *Journal of statistical software*, 45:1–67, 2011.

[Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, et al. Attention is all you need. In *NeurIPS*, 2017.

[Wang *et al.*, 2023] Xu Wang, Hongbo Zhang, Pengkun Wang, Yudong Zhang, Binwu Wang, Zhengyang Zhou, and Yang Wang. An observed value consistent diffusion model for imputing missing values in multivariate time series. In *SIGKDD*, 2023.

[Wang *et al.*, 2024] Shiyu Wang, Jiawei Li, Xiaoming Shi, Zhou Ye, Baichuan Mo, Wenze Lin, Shengtong Ju, Zhixuan Chu, and Ming Jin. Timemixer++: A general time series pattern machine for universal predictive analysis. *arXiv preprint arXiv:2410.16032*, 2024.

[Wen *et al.*, 2023] Qingsong Wen, Tian Zhou, Chaoli Zhang, Weiqi Chen, Ziqing Ma, Junchi Yan, and Liang Sun. Transformers in time series: A survey. In *International Joint Conference on Artificial Intelligence(IJCAI)*, 2023.

[Wu *et al.*, 2023a] Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis. In *ICLR*, 2023.

[Wu *et al.*, 2023b] Yangyang Wu, Jun Wang, Xiaoye Miao, Wenjia Wang, and Jianwei Yin. Differentiable and scalable generative adversarial models for data imputation. *IEEE Transactions on Knowledge and Data Engineering*, 2023.

[Yang *et al.*, 2024a] Xinyu Yang, Yu Sun, Xiaojie Yuan, and Xinyang Chen. Frequency-aware generative models for multivariate time series imputation. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

[Yang *et al.*, 2024b] Yiyuan Yang, Ming Jin, Haomin Wen, Chaoli Zhang, Yuxuan Liang, Lintao Ma, Yi Wang, Chenghao Liu, Bin Yang, Zenglin Xu, et al. A survey on diffusion models for time series and spatio-temporal data. *arXiv preprint arXiv:2404.18886*, 2024.

[Yoon *et al.*, 2019] Jinsung Yoon, William R. Zame, and Mihaela van der Schaar. Estimating missing data in temporal data streams using multi-directional recurrent neural networks. *IEEE Trans. on Biomedical Engineering*, 2019.

[Yuan and Qiao, 2024] Xinyu Yuan and Yan Qiao. Diffusion-ts: Interpretable diffusion for general time series generation. *arXiv preprint arXiv:2403.01742*, 2024.

[Zhou *et al.*, 2023] 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.

[Zhou *et al.*, 2024] Jianping Zhou, Junhao Li, Guanjie Zheng, Xinbing Wang, and Chenghu Zhou. Mtsci: A conditional diffusion model for multivariate time series consistent imputation. In *CIKM*, pages 3474–3483, 2024.