Deep Learning for Time Series Forecasting: A Survey

Xiangjie Kong[1], Zhenghao Chen[1], Weiyao Liu[1], Kaili Ning[1], Lechao Zhang[1], Syauqie Muhammad Marier[1], Yichen Liu[1], Yuhao Chen[1], Feng Xia[2*]

[1]College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, 310023, China.

[2*]School of Computing Technologies, RMIT University, Melbourne, 3000, Australia.

*Corresponding author(s). E-mail(s): f.xia@ieee.org;
Contributing authors: xjkong@ieee.org; 202006010504@zjut.edu.cn; 211123120037@zjut.edu.cn; 211124120018@zjut.edu.cn; 202003151126@zjut.edu.cn; syauqie.mm@gmail.com; liuien@outlook.com; yuhaochen4859@gmail.com;

Abstract

Time series forecasting (TSF) has long been a crucial task in both industry and daily life. Most classical statistical models may have certain limitations when applied to practical scenarios in fields such as energy, healthcare, traffic, meteorology, and economics, especially when high accuracy is required. With the continuous development of deep learning, numerous new models have emerged in the field of time series forecasting in recent years. However, existing surveys have not provided a unified summary of the wide range of model architectures in this field, nor have they given detailed summaries of works in feature extraction and datasets. To address this gap, in this review, we comprehensively study the previous works and summarize the general paradigms of Deep Time Series Forecasting (DTSF) in terms of model architectures. Besides, we take an innovative approach by focusing on the composition of time series and systematically explain important feature extraction methods. Additionally, we provide an overall compilation of datasets from various domains in existing works. Finally, we systematically emphasize the significant challenges faced and future research directions in this field.

时间序列预测（TSF）长期以来是工业界与日常生活中的关键任务。传统统计模型在能源、医疗、交通、气象和经济等领域的实际应用场景中，特别是对预测精度要求较高时，往往存在一定局限性。随着深度学习的持续发展，近年来时间序列预测领域涌现出大量新型模型。然而现有综述研究尚未对该领域广泛采用的模型架构进行统一归纳，亦未对特征提取方法与数据集相关研究作出系统梳理。为填补这一空白，本综述通过全面研究已有成果，从模型架构角度总结了深度时间序列预测（DTSF）的通用范式；创新性地聚焦时间序列的构成要素，系统阐释了重要特征提取方法；同时对现有研究中涉及的多领域数据集进行了整体汇编。最后，我们系统性地强调了该领域面临的重大挑战与未来研究方向

Keywords: Time Series Forecasting, Model Architecture Paradigm, Feature Extraction Methodology, Multivariate Time Series Data

关键词：时间序列预测、模型架构范式、特征提取方法、多元时间序列数据

1 Introduction

Time series are pervasive in various facets of our manufacture and life, serving as a primary dimension to record historical events. Forecasting, a critical task, leverages historical information within sequences to infer the future (Hyndman and Athanasopoulos, 2018; Petropoulos et al., 2022). It finds extensive applications in various domains closely intertwined with our lives, including energy production and consumption (Deb et al., 2017; Li et al., 2019b; Saxena et al., 2019; Rajagukguk et al., 2020; Zhao et al., 2016; Toubeau et al., 2018), meteorological variations (Mudelsee, 2019; Zaini et al., 2022), finance, stock markets, and econometrics (Sezer et al., 2020; Chen and Chen, 2015; Callot et al., 2017; Andersen et al., 2005; Luo et al., 2018; Kalra et al., 2024; Singh et al., 2023) sales and demand (Böse et al., 2017; Bandara et al., 2019) urban traffic flows (Tedjopurnomo et al., 2020; Lv et al., 2014; Li et al., 2017), and welfare-related healthcare conditions (Piccialli et al., 2021; Kaushik et al., 2020; Topol, 2019; Mishra et al., 2024).

时间序列广泛存在于人类生产生活的各个领域，是记录历史事件的重要维度。预测作为其核心任务，旨在利用序列中的历史信息推演未来趋势（Hyndman与Athanasopoulos，2018；Petropoulos等，2022）。该技术在诸多与人类生活密切相关的领域具有重要应用价值，包括能源生产与消费（Deb等，2017；Li等，2019b；Saxena等，2019；Rajagukguk等，2020；Zhao等，2016；Toubeau等，2018）、气象变化（Mudelsee，2019；Zaini等，2022）、金融证券与计量经济学（Sezer等，2020；Chen与Chen，2015；Callot等，2017；Andersen等，2005；Luo等，2018；Kalra等，2024；Singh等，2023）、销售与需求预测（B¨ose等，2017；Bandara等，2019）、城市交通流量（Tedjopurnomo等，2020；Lv等，2014；Li等，2017）以及健康医疗等民生领域（Piccialli等，2021；Kaushik等，2020；Topol，2019；Mishra等，2024）。

Machine learning, data science, and other research groups employing operations research and statistical methods have extensively explored time series forecasting (Fuller, 2009; Faloutsos et al., 2018, 2019b,a). Statistical models typically consider non-stationarity, linear relationships, and specific probability distributions to infer future trends based on the statistical properties of historical data such as mean, variance, and autocorrelation. On the other hand, machine learning models learn patterns and rules from the data. With the emergence (Rosenblatt, 1957) and rapid development of deep learning (Goodfellow et al., 2016; LeCun et al., 2015), an increasing number of neural network models are being applied to time series forecasting. In contrast to the first two approaches that rely on domain-specific knowledge or meaningful feature engineering, deep learning autonomously extracts intricate time features and patterns from complex data. This capability enables the capture of long-term dependencies and complex relationships, ultimately enhancing prediction accuracy. In this article, we will refer to works on Deep Learning for Time Series Forecasting as DTSF works, and Time Series Forecasting will be abbreviated as TSF.

机器学习、数据科学以及运用运筹学与统计方法的研究团队已对时间序列预测领域进行了广泛探索（Fuller, 2009; Faloutsos等, 2018, 2019b,a）。统计模型通常通过考量非平稳性、线性关系及特定概率分布，依据历史数据的均值、方差和自相关等统计特性来推断未来趋势。而机器学习模型则直接从数据中学习规律与规则。随着深度学习技术的兴起（Rosenblatt, 1957）与快速发展（Goodfellow等, 2016; LeCun等, 2015），越来越多的神经网络模型被应用于时间序列预测。相较于前两种依赖领域专业知识或显式特征工程的方法，深度学习能够自主从复杂数据中提取精细的时间特征与模式。这种特性使其能够捕捉长期依赖关系和复杂关联，最终提升预测精度。本文中，我们将时间序列预测深度学习相关研究简称为DTSF研究，时间序列预测则缩写为TSF。

In recent years, deep learning methods have continuously advanced and innovated in time series forecasting (TSF) across various domains (Salinas et al., 2020; Xu et al., 2016; Lai et al., 2018; Bandara et al., 2020; Oord et al., 2016; Rasul et al., 2021; Lim et al., 2021). However, current research efforts primarily focus on key TSF concepts and fundamental model components, while lacking a high-level categorization of deep learning-based DTSF model structures, comprehensive summaries of recent developments, and in-depth analyses of future prospects and challenges. This article aims to address these gaps by drawing on the latest research. The main contributions of this work are as follows:

近年来，深度学习方法在时间序列预测（TSF）领域持续取得突破与创新（Salinas等，2020；Xu等，2016；Lai等，2018；Bandara等，2020；Oord等，2016；Rasul等，2021；Lim等，2021）。然而现有研究主要聚焦于关键TSF概念与基础模型组件，缺乏对基于深度学习的DTSF模型结构的高层次分类、最新进展的系统性总结，以及未来前景与挑战的深度剖析。本文旨在借鉴最新研究成果填补这些空白，主要贡献如下：

- **Dynamic and systematic taxonomy.** We propose a novel dynamic classification method designed to categorize deep learning models for time series forecasting in a systematic manner. Our survey classifies and summarizes these models from the perspective of their architectural structure. To the best of our knowledge, this represents the first dynamic classification of deep learning model architectures for time series forecasting.

动态系统化分类体系。我们提出了一种创新的动态分类方法，旨在以系统化方式对时间序列预测的深度学习模型进行归类。本综述从模型架构结构的视角出发，对这些模型进行了分类与总结。据我们所知，这是首次针对时间序列预测的深度学习模型架构提出的动态分类体系

- **Comprehensive review of data feature enhancement.** We analyze and summarize feature enhancement methods for time series data, including dimensional decomposition, time-frequency transformation, pre-training, and patch-based segmentation. Our analysis begins with the composition of complex, high-dimensional data features, aiming to reveal the latent learning potential within time series data.

数据特征增强的全面综述。我们分析并总结了针对时间序列数据的特征增强方法，包括维度分解、时频变换、预训练以及基于分块的片段化处理。本分析从复杂高维数据特征的构成入手，旨在揭示时间序列数据中潜在的学习能力

- **Summary of challenges and future directions.** This survey summarizes major TSF datasets from recent years, discusses key challenges, and highlights promising future research directions to advance the field.

挑战与未来方向综述。本综述总结了近年来重要的时序预测数据集，探讨了该领域面临的关键挑战，并重点指出了推动学科发展的若干前瞻性研究方向

The remaining content is organized as follows. Section 2 introduces the fundamental aspects of TSF, encompassing the definition and composition of time series, forecasting tasks, statistical models, and existing problems. Section 3, a pivotal component of this paper, mainly delineates the overarching structural paradigm of DTSF models. Section 4 outlines the prevalent paradigms for extracting and learning features from time series data, constituting the second major focus. Section 5 is another key focus of this paper. We not only highlight the limitations and challenges within the current achievements in DTSF research but also elucidate prospective avenues for future exploration. Finally, we conclude this survey in Section 6. In Appendix A, an exhaustive account of TSF datasets across various domains is presented. Figure 1 shows an outline of the entire paper.

本文剩余内容安排如下：第二节阐述时间序列预测（TSF）的基础要素，包括时间序列的定义与构成、预测任务分类、统计模型框架及现存问题。第三节作为全文核心章节，系统阐述深度时间序列预测（DTSF）模型的总体架构范式。第四节详述时间序列数据特征提取与学习的通用范式，构成第二大研究重点。第五节是本文另一关键焦点，不仅剖析当前DTSF研究存在的局限性与挑战，同时展望未来研究的潜在发展方向。第六节为全文总结。附录A完整呈现跨领域TSF数据集详表，图1展示了全文的总体框架结构

2 Time Series Forecasting

Time series represents a continuous collection of data points recorded at regular or irregular time intervals, offering a chronological record of observed phenomena such as vital signs, sales trends, stock market prices, weather changes, and more. The nature of these observations can encompass numerical values, labels, etc. Moreover, time series can be either discrete or continuous (Hamilton, 2020). It is commonly employed for the analysis and prediction of trends and patterns (Montgomery et al., 2015) that evolve over time.

时间序列是指在规则或不规则时间间隔内记录的连续数据点集合，为生命体征、销售趋势、股票价格、天气变化等观测现象提供按时间顺序排列的记录。这些观测值的性质可包含数值型数据、标签型数据等。此外，时间序列既可以是离散的，也可以是连续的（Hamilton, 2020）。该方法通常用于分析和预测随时间演变的趋势与模式（Montgomery等，2015）

TSF is the process of forecasting future values based on the inherent properties and characteristic patterns found in historical data. These properties and intrinsic patterns may provide valuable insights into describing future occurrences. Discovering potential features within time series data based on the similarity of statistical characteristics between adjacent data points or time steps is crucial for building a strong foundation for designing prediction models and achieving improved results.

时间序列预测（TSF）是指基于历史数据中固有的属性和特征模式来预测未来值的过程。这些属性和内在模式可为描述未来事件提供有价值的洞见。根据相邻数据点或时间步之间统计特征的相似性，发掘时间序列数据中的潜在特征，对于构建稳健的预测模型设计基础及获得优化结果具有关键意义

In this section, we will begin with the definition of time series and explain the concept of TSF. Furthermore, we will introduce classical methods based on mathematical

在本节中，我们将从时间序列的定义开始，并解释TSF的概念。此外，我们将介绍基于数理统计的经典方法。最后，我们将分析导致预测精度较低的因素，为该领域的新研究人员提供初步的了解

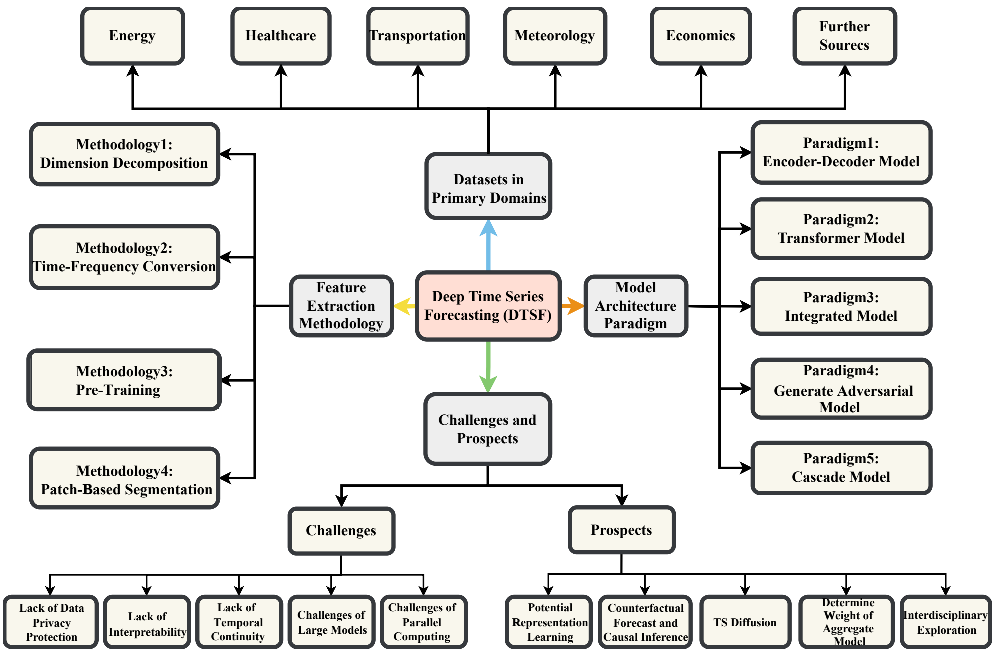


Fig. 1 The outline of this article

statistics. Lastly, we will analyze the factors contributing to lower prediction accuracy to provide researchers new to this field with a preliminary understanding.

2.1 Time Series Definition

在本综述中，我们将时间序列视为按时间顺序记录的观测序列，其观测点之间的时间间隔可以是固定或可变的。设t表示观测时间，yt代表时间序列，对应于随时间观测到的随机变量所构成的随机过程。在大多数情况下，t∈Z，其中Z = (0, ±1, ±2, ...)表示正负整数集合（Fuller, 2009）。当仅能获取有限数据时，时间序列可表示为(y1, y2, y3, ...)。令Y = {yi,1:Ti } N i=1表示N条单变量时间序列的集合，其中yi,1:Ti = (yi,1, ..., yi,Ti)，而yi,t代表第i条时间序列在t时刻的取值。Yt1:t2则是所有N条时间序列在时间区间[t1, t2]内取值的集合

In this survey, we consider time series as observation sequences recorded in chronological order, which may have fixed or variable time intervals between observations. Let t denote the time of observation, and $\mathbf{y}_t$ represents the time series, corresponding to a stochastic process composed of random variables observed over time. In most cases, $t \in \mathbb{Z}$, where $\mathbb{Z} = (0, \pm1, \pm2, \ldots)$ represents the set of positive and negative integers (Fuller, 2009). When only a limited amount of data is available, a time series can be represented as $(\mathrm{y}_1, \mathrm{y}_2, \mathrm{y}_3, \ldots)$. Let $\mathcal{Y} = \{\mathbf{y}_{i,1:T_i}\}_{i=1}^{N}$ denote the collection of N univariate time series, where $\mathbf{y}_{i,1:T_i} = (y_{i,1}, \ldots, y_{i,T_i})$, and $y_{i,t}$ represents the values of t for the i-th time series. $\mathbf{Y}_{t_1:t_2}$ is the collection of values for all N time series within the time interval $[t_1, t_2]$.

时间序列数据与其他类型的数据存在显著差异，因其普遍存在于所有主要领域，并作为构成现实世界的重要维度之一而具有特殊意义。这类数据具有广泛的属性和特征：首先，时间序列数据通常具有高噪声和高维特性，可采用降维技术、小波分析或滤波等方法消除部分噪声并降低维度（Zebari等，2020）；其次，采样时间间隔对其具有重要影响，由于现实世界中固有的不稳定性，不同采样频率下获得的时间序列分布并不具备统一的概率分布特征（Yalavarthi等，2024）

Time series data differs from other forms of data since it is prevalent in all major fields and is significant as one of the aspects that make up our reality. It has a wide range of attributes and characteristics. First of all, time series data are usually noisy and high-dimensional. Techniques such as dimensionality reduction, wavelet analysis, or filtering can be used to eliminate some noise and reduce dimensionality (Zebari et al., 2020). Secondly, the sample time interval has an impact on it. Due to its inherent instability in reality, the distribution of time series obtained at different sampling frequencies does not have a uniform probability distribution (Yalavarthi et al., 2024).

Finally, if time series data is viewed as an information network, each time point can be considered a node, with the relationships between nodes evolving over time. Similar to most real-world networks, this data is inherently heterogeneous and dynamic (Peng et al., 2021), which presents significant challenges for the modeling and analysis of spatio-temporal data. It is worth noting that the representation of time series data is crucial for relevant features extraction and dimensionality reduction. The success or failure of model design and application is closely tied to this representation.

最后，若将时间序列数据视为信息网络，每个时间点可被视为一个节点，节点间的关系随时间动态演化。与大多数现实世界网络类似，此类数据本质上具有异质性和动态性（Peng等, 2021），这为时空数据的建模与分析带来了重大挑战。值得注意的是，时间序列数据的表征对于相关特征提取与降维至关重要，模型设计与应用的成功与否与此表征密切相关

2.2 Forecasting Task

TSF is a process of predicting future data based on historical observations, widely applied in various domains such as energy, finance, and meteorology to anticipate future trends. The task of TSF can be categorized into short-term and long-term forecasting based on the prediction horizon, which is determined by specific application requirements and domain characteristics. Short-term forecasting typically involves shorter time spans, often ranging from hours to weeks, emphasizing high prediction accuracy and is suitable for tasks demanding precision. In contrast, long-term forecasting spans longer periods, including months, years, or even longer durations, and addresses challenges related to long-term trends and seasonal variations that can significantly impact prediction accuracy. The distinction between these two types of forecasting lies in their specific emphasis. Short-term forecasting prioritizes precision and relies mainly on extrapolating data, suitable for scenarios where fluctuations within relatively short periods are critical for prediction outcomes. Conversely, long-term forecasting requires consideration of long-term trends and seasonal influences, making it more complex and necessitating additional factors such as extra assumptions and supplemental external data, which may affect its accuracy. Therefore, the role of external factors is particularly important in long-term forecasting, as they help the forecasting model better capture long-term trends, cyclical fluctuations, and other macro-level changes. For example, external factors such as weather, holidays, economic indicators, and road network information often have a significant impact on the trends and seasonal variations in time series data. Currently, many researchers have incorporated these external factors into forecasting models to improve the accuracy of predictions. Common approaches to handling external influences include incorporating external data as additional features into the model, using multi-task learning with external data (Ruder, 2017), and introducing exogenous variables into classical time series models. Deep learning methods, such as LSTM, GRU, and attention mechanisms, also enhance model performance by considering external factors (Qin et al., 2017). Additionally, seasonal adjustment, periodic modeling, and the integration of road network knowledge are effective methods for addressing external influences. For instance, MultiSPANS (Zou et al., 2024) uses a structural entropy minimization algorithm to generate optimal road network hierarchies, considering complex multi-distance dependencies in the road network for prediction; (Kong et al., 2024), in summarizing forecasting tasks, constructed a new bus station distance network to account for the relationships between external bus stations.

时间序列预测（TSF）是一种基于历史观测数据预测未来值的过程，其广泛应用于能源、金融、气象等诸多领域以实现趋势预判。根据预测周期的不同，TSF任务可分为短期预测与长期预测，这种划分取决于具体应用需求与领域特性。短期预测通常覆盖较短时间跨度（数小时至数周），强调预测精度的高要求，适用于对准确性敏感的预测场景；而长期预测则涉及更长时间范围（数月、数年甚至更长），需要应对可能显著影响预测准确性的长期趋势与季节性变化等挑战。 两类预测的核心差异体现在关注重点上：短期预测以精度为优先，主要依赖数据外推法，适用于较短周期内波动对预测结果起决定性作用的场景；长期预测则需综合考量长期趋势与季节性影响，其复杂性更高，往往需要引入额外假设、补充外部数据等附加因素，这些因素可能对预测准确性产生影响。因此，外部因素在长期预测中具有特殊重要性，它们能帮助预测模型更好地捕捉长期趋势、周期波动等宏观变化。例如天气、节假日、经济指标、路网信息等外部因素，常对时间序列数据的趋势性与季节性变化产生显著影响。目前许多研究者已将此类外部因素纳入预测模型以提升精度，常见处理方法包括：将外部数据作为附加特征输入模型、采用外部数据驱动的多任务学习框架（Ruder, 2017）、在经典时间序列模型中引入外生变量等。深度学习方法（如LSTM、GRU和注意力机制）也通过融合外部因素来增强模型性能（Qin等, 2017）。此外，季节性调整、周期性建模以及路网知识融合均是应对外部影响的有效方法。 典型案例如：MultiSPANS（Zou等, 2024）采用结构熵最小化算法生成最优路网层级结构，通过考虑路网中复杂的多距离依赖关系进行预测；Kong等（2024）在总结预测任务时，构建了新型公交站点距离网络以刻画外部公交站点间的关联关系

On the other hand, in addition to being categorized as Univariate (Zhang et al., 1998; Januschowski et al., 2020; Montero-Manso and Hyndman, 2021; Semenoglou

另一方面，除根据是否考虑多变量将时间序列预测划分为单变量预测（Zhang等，1998；Januschowski等，2020；Montero-Manso和Hyndman，2021；Semenoglou等，2021）与多变量预测（Lütkepohl，2005；Kolassa，2020）外，还可通过全局模型与局部模型的差异进行区分。单变量预测指预测过程中仅考虑单一变量的任务，其核心在于对单个变量未来值的预测；而多变量预测则需同时预测多个相关变量，通过考量变量间的相互依存关系来推断其未来值。 在讨论单变量与多变量预测时，必须关注全局模型与局部模型的区别，这一差异直接影响建模方法及结果解读。全局模型会考虑整个时间序列数据集中的所有变量，而局部模型则聚焦于数据的特定子集（如特定区段或时间窗口），这种差异决定了数据内部依存关系的捕捉方式及预测结果的生成机制

et al., 2021) and Multivariate (Lütkepohl, 2005; Kolassa, 2020) forecasting based on whether multiple variables are considered, TSF can also be distinguished by the distinction between global and local models. Univariate forecasting involves tasks where only one variable is considered during the forecasting process, primarily focusing on predicting the future values of a single variable. Multivariate forecasting, on the other hand, entails the simultaneous prediction of multiple correlated variables, considering the interdependencies among various variables and forecasting their future values. When discussing univariate and multivariate forecasting, it's essential to consider the distinction between global and local models, which impacts the modeling approach and the interpretation of results. Global models consider all variables across the entire time series dataset, while local models focus on subsets of the data, such as specific segments or windows, affecting how dependencies within the data are captured and predictions are made.

综上所述，预测任务的分类与侧重点取决于应用场景及需求。以金融领域为例，短期预测可能涉及对股价分钟级或小时级波动的预测，而长期预测则可能涵盖数周乃至数月的趋势研判。同理在气象学领域，短期预测或需关注数小时内的天气变化，长期预测则可能针对数日乃至数周的天气系统演变。单变量预测可聚焦于特定产品的销量或单一股票的价格走势，而多变量预测则需同步推演多品类产品的销售态势或不同金融市场间的联动关系

In summary, the categorization and focus of forecasting tasks depend on the application context and requirements. For instance, in the financial domain, short-term forecasting may involve predicting stock price fluctuations within minutes or hours, while long-term forecasting could encompass forecasts over several weeks or months. Similarly, in meteorology, short-term forecasting might entail predicting weather conditions within a few hours, while long-term forecasting may involve predictions spanning days or weeks. For univariate forecasting, the focus could be on forecasting the sales volume of a particular product or the price of a specific stock. On the other hand, multivariate forecasting might simultaneously predict the sales volumes of multiple products or the interrelationships within various financial markets.

在后续章节中，我们将介绍统计预测模型并重点阐述其局限性，着重探讨传统时间序列预测方法所面临的挑战。随后，我们将深入分析深度学习预测模型与方法的发展历程

In the following subsections, we will introduce statistical forecasting models and highlight their limitations, emphasizing the challenges posed by traditional TSF methods. Subsequently, we will delve into the development of deep learning forecasting models and methods.

2.3 Statistical Forecasting Model

统计预测模型的发展历程可追溯至20世纪初。如公式1与公式2所示，移动平均法（MA）（Box等，2015；Hipel与McLeod，1994；Cochrane，1997）和简单指数平滑法（ES）（Gardner Jr，1985）等最早的统计预测方法均建立在时间序列基础之上

The development history of statistical forecasting models can be traced back to the early 20th century. Equations 1 and 2 illustrate how the first statistical forecasting methods, such as Moving Averages (MA) (Box et al., 2015; Hipel and McLeod, 1994; Cochrane, 1997) and simple Exponential Smoothing (ES) (Gardner Jr, 1985), were based on time series.

$$MA_t(n) = \frac{1}{n} \sum_{i=t-n+1}^{t} x_i \tag{1}$$

其中，n为窗口大小，MA表示时间t处的移动平均值

where n is the window size, and MA represents the moving average at time t.

$$ES_{t+1} = \alpha \cdot x_t + (1 - \alpha) \cdot ES_t \tag{2}$$

其中，ESt+1表示预测趋势，α为平滑系数，ESt代表上一时间步的预测值。移动平均法通过计算特定时间段内观测值的平均值来平滑数据，而指数平滑法则对近期观测值赋予更高权重，以反映数据的趋势特征

where ES_{t+1} represents the predicted trend, α is the smoothing coefficient, and ES_t is the value predicted at the previous time step. Moving average smooths data by calculating the average of observed values over a certain period of time, while exponential smoothing assigns higher weights to more recent observations to reflect the trend of the data.

Subsequently, the autoregressive (AR) (Box et al., 2015; Hipel and McLeod, 1994; Lee, 1994) and Moving Average (MA) models (represented by Equations 3 and 4, respectively) were introduced as two important concepts, leading to the development of the Autoregressive Moving Average Model (Box et al., 2015; Hipel and McLeod, 1994; Adhikari and Agrawal, 2013) (ARMA, as shown in equation 5). These models aim to accurately capture the auto correlation and averaging properties of time series data.

随后，自回归模型（Autoregressive, AR）（Box等，2015；Hipel和McLeod，1994；Lee，1994）与移动平均模型（Moving Average, MA）（分别由公式3和公式4表示）作为两个重要概念被提出，由此发展出自回归移动平均模型（Box等，2015；Hipel和McLeod，1994；Adhikari和Agrawal，2013）（如公式5所示的ARMA模型）。这些模型旨在精确捕捉时间序列数据的自相关特性与平均特性

$$AR\colon Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \xi_t \tag{3}$$

$$MA\colon Y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} \tag{4}$$

$$\begin{aligned} Y_t = {} & c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} \\ & + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t \end{aligned} \tag{5}$$

where Y_t represents the time series data under consideration, φ_1 to φ_p are parameters of the AR model. These parameters describe the relationship between the current value and values from the past p time points. Similarly, θ_1 to θ_q are parameters of the MA model, which describe the relationship between the current value and errors from the past q time points. ε_t represents the error term at time t, and c denotes a constant term.

其中，Yt表示所研究的时间序列数据，φ1至φp为自回归模型（AR）的参数。这些参数描述了当前值与过去p个时间点数值之间的关系。类似地，θ1至θq为移动平均模型（MA）的参数，用以表征当前值与过去q个时间点误差项之间的关联。εt代表第t时刻的误差项，c表示常数项

Specifically, the AR model leverages past time series observations to predict future values, while the MA model relies on the moving average of observations to make these predictions. To address non-stationary time series data, the Autoregressive Integrated Moving Average (ARIMA) model (Box et al., 2015; Hipel and McLeod, 1994; Cochrane, 1997; Hamzaçebi, 2008; Zhang, 2003) is introduced. ARIMA is employed to transform non-stationary sequences into stationary ones by means of differencing, thereby reducing or eliminating trends and seasonal variations in the time series. This transformation is represented by Equation (6) as follows:

具体而言，自回归（AR）模型利用时间序列的历史观测值预测未来数值，而移动平均（MA）模型则基于观测值的移动平均值进行预测。针对非平稳时间序列数据，研究者提出了自回归积分滑动平均（ARIMA）模型（Box等，2015；Hipel与McLeod，1994；Cochrane，1997；Hamzaçebi，2008；Zhang，2003）。该模型通过差分运算将非平稳序列转化为平稳序列，从而有效削弱或消除时间序列中的趋势项与季节波动。这一转换过程可由公式(6)表示如下：

$$\Delta Y_t = (1 - L)^d Y_t = \epsilon_t \tag{6}$$

where L denotes the lag operator, d represents the differencing order, y_t signifies the time series, and ϵ_t is the error term. This integration of ARIMA helps mitigate non-stationarity, paving the way for more effective TSF.

其中，\(L \) 表示滞后算子，\(d \) 代表差分阶数，\(y_t \) 指代时间序列，\(\epsilon_t \) 为误差项。通过ARIMA模型的这种积分处理，能够有效缓解非平稳性问题，从而为更高效的时间序列预测（TSF）奠定基础

Machine learning models represented by Random Forests and Decision Trees (Rokach, 2016; Ali et al., 2012; Ho, 1995; Kontschieder et al., 2015) offer enhanced flexibility and predictive performance in statistical forecasting (Harvey, 1990; Ahmed et al., 2010). A decision tree comprises a series of decision nodes and leaf nodes, constructed based on the selection of optimal features and splitting criteria to minimize prediction errors or maximize metrics like information gain or Gini index. Each decision node splits based on feature conditions, while each leaf node provides prediction results. Random Forest, on the other hand, makes forecasting by constructing multiple decision trees and combining their forecasting results. It can handle high-dimensional features and large-scale datasets, capturing nonlinear relationships and interactions between features.

以随机森林和决策树为代表的机器学习模型（Rokach，2016；Ali等，2012；Ho，1995；Kontschieder等，2015）在统计预测领域展现出卓越的灵活性和预测性能（Harvey，1990；Ahmed等，2010）。决策树由一系列决策节点和叶节点构成，其构建过程基于最优特征与分割准则的选择，旨在最小化预测误差或最大化信息增益、基尼指数等指标。每个决策节点依据特征条件进行数据划分，而每个叶节点则输出预测结果。随机森林则通过构建多棵决策树并集成其预测结果进行推断，该算法能够有效处理高维特征和大规模数据集，并能捕捉特征间的非线性关系与交互作用

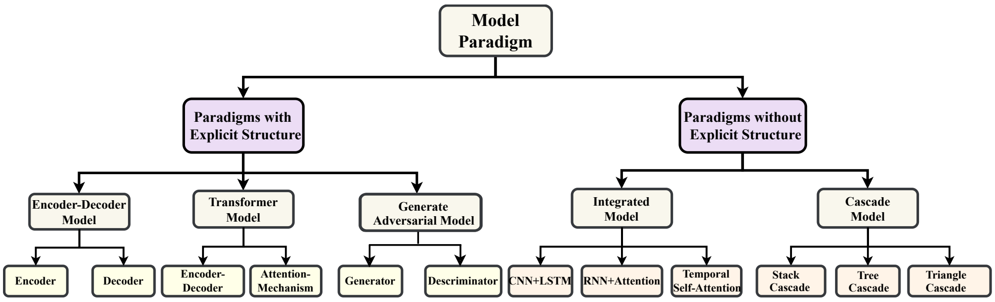


Fig. 2 The details of five paradigms

然而，物联网等新兴技术的发展为数据获取、采集与存储带来了效率提升和便利性（Li等，2015；Kong等，2022）。大数据时代已然来临（Sagiroglu和Sinanc，2013；Fan等，2014），数据正以日益增长的速率持续生成。统计预测模型需要更好地适应处理大规模高维数据的需求（Che等，2013；Wu等，2013；Oussous等，2018）。不同行业与领域对精准预测模型的需求也与日俱增，以支撑决策制定与规划实施（Rodríguez-Mazahua等，2016）。此外，实际应用中还面临着数据间更为复杂的关联关系，这要求模型具备更强的灵活性与精确性来应对这些

However, the development of emerging technologies such as the Internet of Things (IoT) has brought efficiency and convenience to data acquisition, collection, and storage (Li et al., 2015; Kong et al., 2022). The era of big data has arrived (Sagiroglu and Sinanc, 2013; Fan et al., 2014), with data being generated at an increasing rate. Statistical forecasting models need to better adapt to the demands of processing large-scale and high-dimensional data (Che et al., 2013; Wu et al., 2013; Oussous et al., 2018). Different industries and domains are also increasingly in need of accurate forecasting models to support decision-making and planning (Rodríguez-Mazahua et al., 2016). Furthermore, more complex relationships among data are encountered in practical applications, requiring more flexible and accurate models to tackle these challenges.

综上所述，传统统计预测模型在计算能力、预测精度及预测时长方面存在明显局限。该类方法在处理非平稳性、非线性关系、噪声干扰及复杂依赖关系时存在显著缺陷，对长期依赖性与多特征预测任务的适应性亦显不足。随着深度学习模型的持续发展与创新，上述局限性已被逐步突破，从而实现了预测性能的显著提升

In summary, traditional statistical forecasting models are limited in terms of computational power, prediction accuracy, and length. There are major shortcomings in statistical forecasting methods in handling non-stationarity, nonlinear relationships, noise, and complex dependencies, and their adaptability to long-term dependencies and multi-feature forecasting tasks is also limited. With the continuous development and innovation of deep learning models, these limitations have been overcome, leading to improved predictive performance.

3 DTSF Model Architecture

时间序列数据普遍存在于能源、交通、通信系统等诸多现实领域。对时间序列数据进行精确建模与预测，对提升这些系统的运行效率具有关键作用。经典深度学习模型（RNN、TCN、Transformer和GAN）在时间序列预测领域取得了显著进展（Wu等，2021；Zhou等，2022b；Woo等，2022b；Zhang等，2022b），为后续研究提供

Time series data is prevalent in various real-world domains, including energy, transportation, and communication systems. Accurately modeling and predicting time series data plays a crucial role in enhancing the efficiency of these systems. Classical deep learning models (RNN, TCN, Transformer, and GAN) have made significant advancements in TSF (Wu et al., 2021; Zhou et al., 2022b; Woo et al., 2022b; Zhang et al., 2022b), providing valuable insights for subsequent research.

广泛采用的方法之一是循环神经网络（RNN），它利用循环连接来处理时间关系并捕获序列数据中的演化模式。RNN的变体，即长短期记忆（LSTM）和门控循环单元（GRU），是专门为解决长期依赖性和有效捕获长时间序列中的模式而设计的。有很多基于RNN的研究，DeepAR（Salinas等人，2020）利用RNN和自回归技术来捕捉时间序列数据中的时间依赖性和模式

One of the widely adopted methods is the Recurrent Neural Network (RNN), which utilizes recurrent connections to handle temporal relationships and capture evolving patterns in sequential data. Variants of RNNs, namely Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), are specifically designed to address long-term dependencies and effectively capture patterns in long time series. There is a lot of research based on RNNs, DeepAR (Salinas et al., 2020) leveraged RNN and autoregressive techniques to capture temporal dependencies and patterns in time series

Table 1 DTSF Model Architecture Paradigm

Architecture	Model	Multi/Uni	Output	Loss	Metrics	Year
Encoder - Decoder	COST (Woo et al., 2022a)	Multi & Uni	Point	contrastive loss	MSE, MAE	2022
	TS2Vec (Yue et al., 2022)	Multi & Uni	Point	contrastive loss	MSE	2022
	ACT (Li et al., 2022c)	Multi & Uni	Point	cross-entropy	Q50 loss, Q90 loss	2022
	SimTS (Zheng et al., 2023)	Multi & Uni	Point	cos-similarity loss, InfoNCE loss	MAE, MSE	2023
	DeepTCN (Chen et al., 2020b)	Multi	Pro	quantile loss	NRMSE, SMAPE, MASE	2020
	STEP (Shao et al., 2022)	Multi	Pro	MAE	MAE, RMSE, MAPE	2022
	DCAN (He et al., 2022)	Multi	Point	RMSE	MAE, RMSE	2022
	FusFormer (Yang and Lu, 2022)	Multi	Point	–	RMSE, RMSE Decrease	2022
	HANet (Bi et al., 2023)	Multi	Point	–	MAE, RMSE	2022
	D^3VAE (Li et al., 2022b)	Multi	Pro	–	MSE, CRPS	2022
	TI-MAE (Li et al., 2023b)	Multi	Point	MSE	MSE, MAE	2023
	AST (Wu et al., 2020b)	Uni	Pro	cross-entropy	Q50, Q90 loss	2020
Transformer	TFT (Lim et al., 2021)	Multi & Uni	Prob	quantile loss	P50, P90 quantile loss	2021
	Informer (Zhou et al., 2021)	Multi & Uni	Point	MSELoss	MSE, MAE	2021
	ETSformer (Woo et al., 2022b)	Multi & Uni	Point	MSELoss	MSE, MAE	2022
	FEDformer (Zhou et al., 2022b)	Multi & Uni	Point	MSELoss	MSE, MAE, Permutation	2022
	TACTiS (Drouin et al., 2022)	Multi & Uni	Pro	log-likelihood	CRPS-Sum, CRPS-means	2022
	Autoformer (Wu et al., 2021)	Multi & Uni	Point	L2 loss	MSE, MAE	2022
	NSTformer (Liu et al., 2022b)	Multi & Uni	Point	L2 loss	MSE, MAE	2023
	Dateformer (Young et al., 2022)	Multi & Uni	Point	MSE	MSE, MAE	2023
	Crossformer (Zhang and Yan, 2023)	Multi & Uni	Point	MSE	MSE, MAE	2023
	Scaleformer (Shabani et al., 2022)	Multi & Uni	Pro	MSE	MSE, MAE	2023
	BasisFormer (Ni et al., 2023)	Multi & Uni	Point	MSE	MSE, MAE	2023
	CRT (Zhang et al., 2022a)	Multi	Point	–	ROC-AUC, F1-Score	2021
	Pyraformer (Liu et al., 2021)	Multi	Point	MSE	MSE, MAE	2022
	TDformer (Zhang et al., 2022b)	Multi	Point	MSE	MSE, MAE	2022
	FusFormer (Yang and Lu, 2022)	Multi	Point	–	RMSE, RMSE Decrease	2022
	Scaleformer (Shabani et al., 2022)	Multi	Point	MSE, Huber, Adaptive loss	MSE, MAE	2022
	Infomaxformer (Tang and Zhang, 2023)	Multi	Pro	MSELoss	MSE, MAE	2023
	PatchTST(Nie et al., 2022)	Multi	Point	Adaptive Loss	MSE, MAE	2023
	iTransformer(Liu et al., 2023c)	Multi	Point	L2 Loss	MSE, MAE	2023
	MCformer (Han et al., 2024)	Multi	Point	MSE, MAE	MSE, MAE	2024
	SAMformer (Ilbert et al., 2024)	Multi	Point	MSE	MSE, MAE	2024
	TSLANet (Eldele et al., 2024)	Multi	Point	MSE	MSE, MAE	2024
	MASTER (Li et al., 2024a)	Multi	Point	MSE	IC, ICIR, RankIC	2024
	TimeSiam (Dong et al., 2024)	Multi	Point	L2, Cross-Entropy	MSE, MAE, Recall, F1 Score	2024
	Chronos (Ansari et al., 2024)	Multi	Point	Cross Entropy	WQL, CRPS, MASE	2024
	TimeXer (Wang et al., 2024c)	Multi	Point	L2 loss	MSE, MAE	2024
	Time-SSM (Hu et al., 2024)	Multi	Point	MSE	MSE, MAE	2024
	SageFormer (Zhang et al., 2024)	Multi	Point	MSE	MSE, MAE	2024
	TIME-LLM (Wang et al., 2024a)	Multi	Point	MSE, SMAPE	MSE, MAE, SMAPE	2024
	CARD (Wang et al., 2024b)	Multi	Point	MSE, MAE	MSE, MAE	2024
	Pathformer (Chen et al., 2024)	Uni	Pro	L1 loss	MSE, MAE	2024
GAN	ForGAN (Koochali et al., 2019)	Multi & Uni	Pro	RMSE	MAE, MAPE, RMSE	2019
	COSCI-GAN (Seyfi et al., 2022)	Multi & Uni	Pro	Global loss = local + central	MAE	2022
	RCGAN (Esteban et al., 2017)	Multi	Pro	cross-entropy	AUROC, AUPRC	2017
	TimeGAN (Yoon et al., 2019)	Multi	Pro	Unsupervised, Supervised, Reconstruction, Loss	Discriminative and Predictive Score	2019
	PSA-GAN (Jeha et al., 2022)	Multi	Point	Wasserstein loss	–	2022
	AEC-GAN (Wang et al., 2023)	Multi	Point	MSE	ACF, Skew / Kurt, FD	2023
	ITF-GAN (Klopries and Schwung, 2024)	Multi	Point	MSE	MSE, STS, Pearson, Hellinger, Pred.	2024
	MAGAN (Ferchichi et al., 2024)	Multi	Point	–	MAE, MAPE	2024
	TSGAN (Xu et al., 2024)	Multi	Point	–	MAE, RMSE, MAPE	2022
	AST (Wu et al., 2020b)	Uni	Pro	cross-entropy	Q50 loss, Q90 loss	2020
Integrated Module	ConvLSTM (Shi et al., 2015)	Multi & Uni	Point	cross-entropy	Rainfall-MSE, CSI, FAR, POD	2015
	Bi-LSTM (Du et al., 2020)	Multi	Point	MSE	MAE, RMSE	2020
	(Fu et al., 2022a)	Multi	Point	MAE	MAE, RMSE, MAPE	2022
	(Asiful et al., 2018)	Uni	Point	L2 loss	MAE, MSE, MAPE	2018
	TATCN (Wang and Zhang, 2022)	Uni	Point	–	MAE, RMSE, sMAPE	2022
	LST-TCN (Sheng et al., 2022)	Uni	Point	Pinball loss	MAPE, RMSE	2022
Cascade	TimesNet (Wu et al., 2022)	Multi & Uni	Point	MSE, SMAPE	MSE, MAE, SMAPE, MASE	2022
	TreeDRNeT (Zhou et al., 2022c)	Multi & Uni	Point	Lp Regularized Loss	MSE, MAE	2022
	Triformer (Cirstea et al., 2022)	Multi & Uni	Point	–	MSE, MAE	2022
	SCINet (Liu et al., 2022a)	Multi & Uni	Point	L1 loss	RSE, CORR, MSE, MAE, MAPE, RMSE	2022
	HTSF (Duan et al., 2023)	Multi	Pro	L2 loss, HyperGRU	MAE, RMSE	2023
	CIPM (Yolcu and Yolcu, 2023)	Multi	Point	–	RMSE, MAPE, MdRAE	2023
	MACN (He et al., 2023)	Multi	Point	RMSE	RMSE, MAE	2023
	CasCIFF (Zhu et al., 2024)	Multi	Point	–	MSLE, MAPE	2024
	FCPM (Guo et al., 2024)	Multi	Point	RMSE	MAE	2024
	N-BEATS (Oreshkin et al., 2019)	Uni	Point	MAE	SMAPE, OWA, MASE	2020

MQRNN（Wen等人，2017年）通过融合循环神经网络的表达力与时序特性、分位数回归的非参数特性以及直接多步预测的高效性，提出了一种名为"分叉序列"的新型训练方案，显著提升了模型的稳定性与预测性能。ES-RNN（Smyl，2020年）则构建了一个动态计算图神经网络框架，将标准指数平滑模型与长短期记忆网络整合于统一

data. MQRNN (Wen et al., 2017) exploited the expressiveness and temporal nature of RNNs, the nonparametric nature of Quantile Regression and the efficiency of Direct Multi Horizon Forecasting, proposed a new training scheme named forking-sequences to boost stability and performance. ES-RNN (Smyl, 2020) proposed a dynamic computational graph neural network with a standard exponential smoothing model and LSTM in a common framework.

除循环神经网络外，卷积神经网络（CNN）同样可应用于时间序列预测任务。通过将时间序列数据视为一维信号进行处理，卷积神经网络能够从局部区域提取特征，从而有效捕捉局部模式并保持平移不变性。值得注意的是，时序卷积网络（TCN）是基于卷积神经网络的时间序列分析模型中具有代表性的范例

In addition to RNNs, Convolutional Neural Networks (CNNs) can also be employed for TSF. By processing time series data as one-dimensional signals, CNNs can extract features from local regions, enabling them to capture local patterns and translational invariance effectively. Notably, Temporal Convolutional Networks (TCNs) represent a prominent example of CNN-based models for time series analysis.

时态卷积网络（Temporal Convolutional Network）是一种经典的深度学习模型，因其能有效捕捉长程依赖关系而在时间序列预测领域广受关注。与传统循环神经网络不同，该模型采用带有空洞卷积的卷积层结构，可在不增加参数量的前提下扩展感受野。这种特性使时态卷积网络在保持计算效率的同时，能更高效地处理长程依赖问题（Bai等，2018a）。对于具有复杂时间模式的序列数据，该模型能有效建模不同长度的时序关系，且不会出现梯度消失问题（Deng等，2019）。在交通流量预测中，该技术已成功应用于传感器数据的时间依赖性建模，实现了高精度的交通状况预测（Zhao等，2019）。值得注意的是，当与注意力机制、特征提取层等技术结合时，时态卷积网络在各类预测任务中展现出更优越的性能。例如，在电力负荷预测、能源需求预测等多变量时间序列预测任务中，基于注意力机制增强的时态卷积网络模型取得了显著提升的预测效果。总体而言，该模型为时间序列预测提供了强大而高效的解决方案，尤其适用于处理长序列或具有复杂时间依赖性的数据

The Temporal Convolutional Network is a classical deep learning model that has garnered widespread attention in time series forecasting due to its ability to effectively capture long-range dependencies. Unlike traditional RNN, TCNs employ convolutional layers with dilated convolutions to expand the receptive field without increasing the number of parameters. This enables TCNs to handle long-range dependencies more efficiently while maintaining computational efficiency (Bai et al., 2018a). TCNs are particularly useful for time series data with complex temporal patterns, as they can model sequences of varying lengths without suffering from the vanishing gradient problem (Deng et al., 2019). In traffic flow prediction, TCNs have been successfully applied to model the temporal dependencies in sensor data, achieving high accuracy in forecasting traffic conditions (Zhao et al., 2019). Furthermore, when combined with other techniques such as attention mechanisms and feature extraction layers, TCNs have demonstrated improved performance across various prediction tasks. For instance, integrating TCNs with attention-based models has shown enhanced results in multivariate time series forecasting tasks like electricity load prediction and energy demand forecasting. Overall, TCNs provide a powerful and effective approach to time series forecasting, especially when dealing with long sequences or datasets with complex temporal dependencies.

另一种有价值的技法是注意力机制，该机制使模型能够对输入序列的不同部分赋予差异化权重。这一特性尤其有利于处理长期序列或在特定时间节点聚焦关键信息。此外，生成对抗网络（GAN）也可应用于时间序列预测领域。通过生成器与判别器之间的对抗训练，生成对抗网络能够合成时间序列样本，从而提供更精确的预测结果

Another valuable technique is the attention mechanism, which allows models to assign varying weights to different parts of the input sequence. This is particularly beneficial for handling long-term series or focusing on important information at specific time points. Additionally, Generative Adversarial Networks (GANs) can be utilized for TSF. Through adversarial training between a generator and a discriminator, GANs can generate synthetic time series samples and provide more accurate prediction results.

在本节中，我们基于模型架构维度对现有时间序列模型进行动态分类。研究聚焦于模型的内部结构设计，将五种模型架构划分为显式结构范式与隐式结构范式两类。图2展示了我们提出的模型分类方案的更多细节。表1全面汇总了近年来做出突出贡献的各类模型，表2则精选若干关键模型，就其优势、缺陷、应用领域及预测范围展开详细解析。此举旨在帮助读者理解各模型的独特性，并为特定预测任务选择最适配模型提供指导

In this section, we dynamically classify existing time series models based on the model architecture dimension. We focus on the internal structural design of the models and categorize the five model architectures into explicit structure paradigms and implicit structure paradigms. Figure 2 shows more details of our proposed model classification. Table 1 comprehensively summarizes the models that have made outstanding contributions in recent years. Table 2 selects several key models and provides a detailed analysis of their advantages, disadvantages, application domains, and prediction horizons. The aim is to help readers understand the unique characteristics of each model and guide them in selecting the most suitable model for specific prediction tasks.

3.1 Model with Explicit Structure

3.1.1 Encoder-Decoder Model

The encoder-decoder model is widely used in the field of deep learning, which appears similar to seq2seq and has an explicit encoder and a decoder. However, seq2seq seems to be described from an application-level perspective, while the encoder-decoder is described at the network level. U-net for medical image segmentation (Ronneberger et al., 2015) and various forms of Transformers are well-known applications.

编码器-解码器模型在深度学习领域应用广泛，其结构与序列到序列（seq2seq）模型相似，均包含显式的编码器和解码器模块。然而，序列到序列模型通常是从应用层面进行描述，而编码器-解码器模型则更多在网络架构层面进行讨论。医学图像分割中的U-Net网络（Ronneberger等，2015）以及各类Transformer变体，都是该模型的典型应用实例

Table 2 A comparative analysis of time series forecasting models: advantages, disadvantages, applications, and prediction lengths. 时间序列预测模型的比较分析：优势、劣势、应用场景及预测长度

Model	Advantages	Disadvantages	Applications	Prediction Horizon
Informer (Zhou et al., 2021)	Effcient; Strong repre--sentation; Good generalization	Sensitive to data shifts	Energy; Weather	Long
HANet (Bi et al., 2023)	Capture complex dependencies; Flexible for multivariate data	High complexity	Weather; Ecology	Long
Autoformer (Wu et al., 2021)	Efficient; Good information	High complexity; Depend on data periodicity	Finance; Energy; Electricity; Traffic; Weather; Healthcare	Long
ETSformer (Woo et al., 2022b)	Combine traditional methods with Transformer; Adaptive time window	High computational cost; Requires large data	Finance; Energy; Electricity; Traffic; Weather; Healthcare	Short
FEDformer (Zhou et al., 2022b)	Frequency enhancement; Better flexibility for long-term forecasts	High complexity; Large data needed	Finance; Energy; Electricity; Traffic; Weather; Healthcare	Long
TreeDRNet (Zhou et al., 2022c)	Capture time dynamics; Efficient training with joint networks	High complexity; Need large data.	Finance; Energy; Electricity; Traffic; Weather; Healthcare	Long
TATCN (Wang and Zhang, 2022)	Capture temporal dependencies; Extract local patterns.	High computational cost; Data dependence.	Electricity; Healthcare	Short

In this context, the classic Seq2Seq model stands as one of the most representative Encoder-Decoder architectures. It uses Long Short-Term Memory networks as both the encoder and decoder to map input sequences to output sequences, making it particularly suitable for multi-step forecasting tasks (Sutskever, 2014). Additionally, LSTM and GRU are classic models for time series data modeling, capable of capturing long-term dependencies, and have demonstrated excellent performance in various time series forecasting tasks, such as financial forecasting and weather prediction (Cho,

在此背景下，经典的Seq2Seq模型是最具代表性的编码器-解码器架构之一。该模型采用长短期记忆网络同时作为编码器和解码器，将输入序列映射至输出序列，使其特别适用于多步预测任务（Sutskever，2014）。此外，LSTM和GRU作为时间序列数据建模的经典模型，能够捕捉长期依赖关系，并在金融预测、气象预报等多种时间序列预测任务中展现出卓越性能（Cho，2014）。与传统RNN不同，时序卷积网络（TCN）通过卷积层处理长期依赖问题，在交通流量预测、气象预报等应用中取得显著效果（Bai等，2018a）。而双向编码器-解码器模型利用双向LSTM同时捕获历史与未来时间信息，进一步提升了模型的预测精度（Cheng等，2022）。这些经典编码器-解码器模型凭借其自动学习时间序列复杂模式的能力，已成为时间序列预测任务中不可或缺的工具

2014). In contrast to traditional RNNs, TCN leverage convolutional layers to address long-term dependency issues, achieving strong results in several time series forecasting applications, particularly in traffic flow prediction and weather forecasting (Bai et al., 2018a). Moreover, the Bi-directional Encoder-Decoder model, which utilizes bidirectional LSTM, captures both past and future time information, further enhancing the model's forecasting accuracy (Cheng et al., 2022). These classic Encoder-Decoder models, with their ability to automatically learn complex patterns in time series data, have become essential tools in time series forecasting tasks.

编码器-解码器架构在时间序列预测（TSF）领域同样得到了广泛且成功的应用。例如，Perslev等人（2019）受U-Net（Ronneberger等，2015）启发，基于U型网络架构设计了一种名为U-Time的全卷积时序网络。该网络能够将任意长度的序列输入映射至用户自定义时间尺度上的标签序列，其整体结构呈现高度对称的编码器-解码器U型拓扑。我们认为这种高度对称性的设计源于网络输入与输出存在于同一空间域的特性：编码器将输入映射至另一空间后，解码器需从该空间进行逆向映射还原，因此从理论层面决定了网络架构必须具备高度对称性

Encoder-decoder has also been extensively and successfully applied in the field of TSF. For instance, Perslev et al. (2019) was inspired by U-net (Ronneberger et al., 2015) and designed a time fully convolutional network called U-Time based on the U-net architecture. U-Time maps arbitrarily long sequential inputs to label sequences on a freely chosen time scale. The overall network exhibits a U-shaped architecture with highly symmetric encoder and decoder components. We believe that the high degree of symmetry in the architecture is because the proposed network's input and output exist in the same space. The encoder maps the input into another space, and the decoder should map back from this space. Therefore, the network architecture is theoretically highly symmetric.

存在诸多高度对称的编码器-解码器网络架构，同时也存在编码器与解码器不对称的案例。最具代表性的当属Transformer架构（Zhou等，2021；Wu等，2021；Zhou等，2022b；Yang与Lu，2022）。可观察到解码器不仅与编码器存在差异，还需接收特定输入。这种编码器-解码器架构被认为需要辅助性额外信息才能实现更优性能

There are many highly symmetric encoder-decoder network architectures, as well as cases where the encoder and decoder are asymmetric. The most typical example is the Transformer architecture (Zhou et al., 2021; Wu et al., 2021; Zhou et al., 2022b; Yang and Lu, 2022). It can be observed that the decoder differs from the encoder and receives input. This encoder-decoder architecture is considered to require additional information for assistance to perform better.

同样地，Guo等学者（2023）提出了一种非对称编码器-解码器学习框架，该框架通过卷积神经网络和门控循环神经网络提取多栋建筑间的空间关系与时间序列特征，从而在编码器中形成新的输入数据。解码器则借助注意力机制基于这些输入数据进行预测

Likewise, Guo et al. (2023) proposed an asymmetric encoder-decoder learning framework where the spatial relationships and time-series features between multiple buildings are extracted by a convolutional neural network and a gated recurrent neural network to form new input data in the encoder. The decoder then makes predictions based on the input data with an attention mechanism.

此处还存在其他编码器-解码器架构的例证。Bi等人（2023年）提出了一种用于多元时间序列长期预测的新型分层注意力网络（HANet），该架构同样包含编码器与解码器模块。然而值得注意的是，其编码器与解码器结构存在显著差异，即二者呈现非对称特性。此外，某些网络架构虽明确包含编码器组件，却未设置显式的解码器模块（Eldele等人，2021年）

There are some other examples of encoder-decoder here as well. In Bi et al. (2023), a novel hierarchical attention network (HANet) for the long-term prediction of multivariate time series was proposed, which also includes an encoder and a decoder. However, the encoder and decoder architectures are noticeably different. That is to say, the encoder and decoder are asymmetric. There are also network architectures that explicitly involve an encoder but lack an explicit decoder(Eldele et al., 2021).

3.1.2 Transformer Model

随着Transformer在计算机视觉和自然语言处理（NLP）领域的卓越表现，该模型也被应用于时间序列预测（TSF）领域并展现出巨大潜力。Transformer的核心架构主要由注意力机制和编码器-解码器结构构成

With the remarkable performance of Transformer in computer vision and Natural Language Processing (NLP) domains, they have also been applied to the field of TSF and have shown great promise. The main architecture of the Transformer includes the attention mechanism and the encoder-decoder architecture.

However, applying Transformer to TSF tasks is not without challenges and limitations. Recent studies have highlighted several issues, such as the inability to directly handle Long Sequence Time Forecasting (LSTF), including quadratic time

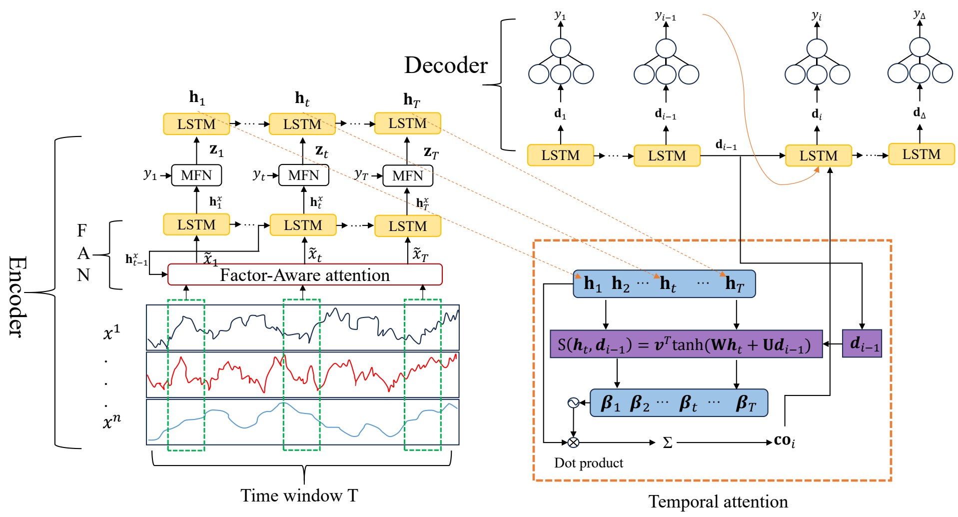


Fig. 3 The overview of HANet model

complexity, high memory usage, and inherent limitations of the encoder-decoder architecture. To address these limitations, Informer (Zhou et al., 2021) was an efficient Transformer-based architecture specifically designed for LSTF. This architecture utilizes the ProbSparse self-attention mechanism, which reduces the time complexity and memory usage to O(LlogL). From the network architecture perspective, it is evident that Informer's architecture(Zhou et al., 2021) closely resembles the vanilla Transformer, consisting of an encoder and a decoder. The encoder receives the input, and the decoder receives the output from the encoder as well as the input, with the addition of zero-padding in the parts to be predicted. The self-attention mechanism is replaced with the ProbSparse self-attention mechanism. TFT (Lim et al., 2021) proposed other architectural improvements to improve accuracy and computational complexity, which integrates high-performance multi-horizon forecasting with interpretable insights into temporal dynamics, capturing temporal relationships at different scales by employing recurrent layers for local processing and interpretable self-attention layers for long-term dependencies.

Autoformer (Wu et al., 2021), on the other hand, argues that previous Transformer-based prediction models (e.g., Informer (Zhou et al., 2021)) mainly focused on improving self-attention for sparse versions. While significant performance improvements were achieved, they sacrificed the utilization of information. One of the reasons why Transformer cannot be directly applied to LSTF is the complex characteristics of time series data. Without special design, traditional attention mechanisms struggle to model and learn these characteristics. Autoformer (Wu et al., 2021) adopts decomposition as a standard approach for time series analysis (Makridakis, 1978; Cleveland et al., 1990), as it is believed that decomposition can untangle the intertwined time patterns and highlight the intrinsic properties of time series. Autoformer (Wu et al., 2021) introduces a novel decomposition architecture with autocorrelation mechanisms,

然而，将 Transformer 应用于时间序列预测（TSF）任务并非没有挑战和局限性。近期研究指出了几个问题，例如其无法直接处理长序列时间预测（LSTF），这包括二次时间计算复杂度、高内存消耗，以及编码器-解码器架构固有的局限性。为应对这些局限性，Informer (Zhou et al., 2021) 提出了一种专为 LSTF 设计的高效基于 Transformer 的架构。该架构采用了 ProbSparse 自注意力机制，将时间复杂度和内存消耗降至 O(L log L)。从网络架构角度看，显然 Informer (Zhou et al., 2021) 的架构非常类似于原始 (vanilla) Transformer，由一个编码器和一个解码器组成。编码器接收输入，解码器则接收来自编码器的输出以及输入本身，并对需要预测的部分进行零填充。其自注意力机制被替换为 ProbSparse 自注意力机制。另一方面，TFT (Lim et al., 2021) 提出了其他架构改进以提高预测精度和降低计算复杂度。它将高性能的多水平预测与对时间动态模式的可解释性洞察相结合，通过采用循环层进行局部处理、并使用可解释的自注意力层捕获长期依赖关系，从而捕捉不同时间尺度上的时序关系

另一方面，Autoformer (Wu et al., 2021) 提出，以往基于 Transformer 的预测模型（例如 Informer (Zhou et al., 2021)）主要集中于改进稀疏化版本的自注意力机制。虽然取得了显著的性能提升，但它们牺牲了信息的利用效率。Transformer 无法直接应用于长序列时间预测 (LSTF) 的原因之一在于时间序列数据的复杂性。若缺乏专门设计，传统的注意力机制难以建模和学习这些特性。因此，Autoformer (Wu et al., 2021) 采用分解作为时间序列分析的标准方法 (Makridakis, 1978; Cleveland et al., 1990)，因为人们认为分解能够理清相互交织的时间模式，并突出时间序列的内在属性。Autoformer (Wu et al., 2021) 引入了一种结合自相关机制的全新分解架构，这与传统的序列分解预处理方法不同。在网络架构方面，其遵循与 Transformer、Informer 及其他类似架构相似的宏观架构。区别在于解码器 (Decoder) 的输入不再是原始输入，而是通过时间序列分解得到的子序列，这些子序列包含季节性和趋势性维度

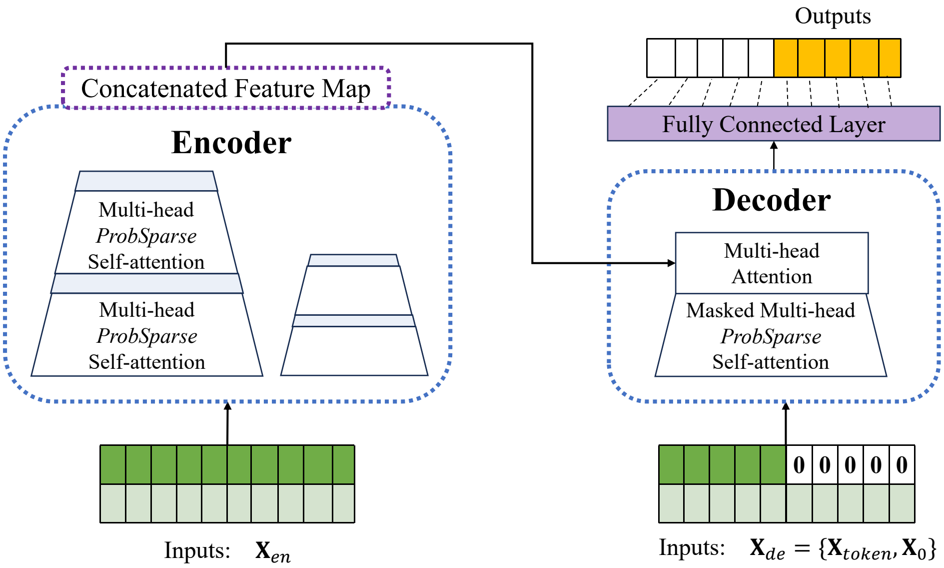


Fig. 4 The overview of Informer model

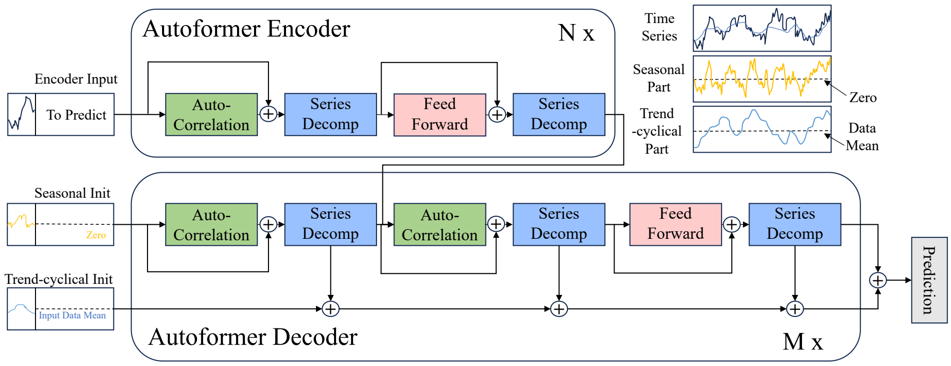


Fig. 5 The overview of Autoformer model

which is different from the conventional series decomposition preprocessing. In terms of the network architecture, it follows a macro architecture similar to Transformers, Informer, and other architectures. The difference lies in the input to the Decoder, which is no longer the original input but rather sub-sequences obtained through time series decomposition, including seasonal and trend dimensions.

In time series forecasting tasks, many researchers prefer to divide long time series into smaller segments to help Transformer models focus more effectively on local temporal features. This approach enhances the model's ability to learn local patterns while

在时间序列预测任务中，许多研究者倾向于将长序列分割为较短的片段，以帮助Transformer模型更有效地聚焦局部时序特征。这种方法在增强模型局部模式学习能力的同时，还能降低计算负担。TSMixer（Ekambaram等人，2023）采用了类似策略，将时序数据划分为多个补丁后，通过基于MLP的层级结构处理这些补丁以提取特征。这种与计算机视觉中基于补丁的方法相仿的处理方式，使模型在时序预测任务中既能有效捕捉局部特征，又可降低计算复杂度和内存需求。Zhang等人（2023b）在其研究MTPNet中提出了一种新型的基于Transformer的多变量时序建模方法，通过对时序分解补丁的季节性成分同时进行时间和空间维度的嵌入，实现了任意粒度时序信息的建模

reducing computational burden. TSMixer (Ekambaram et al., 2023) adopts a similar strategy by partitioning time series data into multiple patches and then processing these patches through MLP-based layers to extract features. This approach, akin to patch-based methods in computer vision, enables the model to capture local features effectively while reducing computational complexity and memory requirements in time series forecasting tasks. Zhang et al. (2023b) proposed a novel Transformer-based multivariate time series modeling approach in their work, MTPNet. It achieves modeling of temporal information at arbitrary granularities by simultaneously embedding temporal and spatial dimensions of the Seasonal part of the time series decomposition patches.

There are further works addressing Transformers in the context of TSF. ETSformer (Woo et al., 2022b) argues that the sequence decomposition used by Autoformer makes simplified assumptions and is insufficient to properly model complex trend patterns. Considering that seasonal patterns are more easily identifiable and detectable, ETSformer designs exponential smoothing attention (ESA) and frequency attention (FA) mechanisms. The network architecture decomposes the time series into interpretable sequence components such as level, growth, and seasonality. FEDformer combines Transformers with seasonal-trend decomposition methods. The decomposition method captures the global profile of the time series, while the Transformer captures more detailed architectures, making it a frequency-enhanced Transformer.

在时间序列预测（TSF）领域，针对Transformer模型的研究持续深入。ETSformer（Woo等人，2022b）指出Autoformer采用的序列分解方法存在简化假设，难以充分建模复杂趋势模式。鉴于季节性模式更易识别和检测，该研究设计了指数平滑注意力（ESA）和频率注意力（FA）机制，其网络架构可将时间序列分解为可解释的组成成分，如水平项、增长项和季节性项。FEDformer则将Transformer与季节性-趋势分解方法相结合：分解模块负责捕捉时间序列的全局轮廓，而Transformer结构则专注于更精细的特征建模，由此构建出频率增强型Transformer架

These studies demonstrate the ongoing efforts in leveraging Transformers for TSF and the development of specialized architectures and mechanisms to overcome the challenges and limitations associated with applying Transformers to this domain.

这些研究表明，学界正持续致力于利用Transformer模型进行时间序列预测，并通过开发专用架构与机制来克服该领域应用Transformer所面临的挑战与局限性

3.1.3 Generative Adversarial Model

GAN (Generative Adversarial Networks) has attracted significant attention since its introduction as a generative model consisting of an explicit structure including a discriminator and a generator. While GANs have been widely used in the field of computer vision, their application in TSF has been relatively limited. The reason for this limited usage is speculated to be the availability of alternative metrics such as CRPS (Continuous Ranked Probability Score) that can measure the quality of generated samples (Benidis et al., 2022).

生成对抗网络（GAN）自提出以来便因其由判别器与生成器构成的显式结构而作为生成模型受到广泛关注。尽管GAN在计算机视觉领域已得到普遍应用，但在时间序列预测（TSF）中的应用仍相对有限。据推测，这种局限性源于存在诸如连续排序概率评分（CRPS）等可衡量生成样本质量的替代性评估指标（Benidis等，2022）

In the existing literature on GAN-based TSF, most studies focus on generating synthetic time series datasets (Yoon et al., 2019; Esteban et al., 2017; Takahashi et al., 2019). The discriminator is trained to distinguish between real and generated time series data, with the goal of producing synthetic data that is indistinguishable from real data. TimeGAN (Yoon et al., 2019), a GAN-based network architecture, was proposed to generate realistic time series data by leveraging the flexibility of unsupervised models and the control of supervised models. It utilizes an embedding function and a recovery function to extract high-dimensional features from time series data, which are then fed into the sequence generator and sequence discriminator for adversarial training. Another study proposed a GAN-based network architecture using Recurrent Neural Networks (RNNs) to generate real-valued multidimensional time series (Takahashi et al., 2019). The study introduced two variations, Recursive GAN (RGAN) and Recursive Conditional GAN (RCGAN), where RGAN generates real-valued data

在现有关于基于生成对抗网络的时间序列生成（GAN-based TSF）的文献中，大多数研究集中于生成合成时间序列数据集（Yoon等，2019；Esteban等，2017；Takahashi等，2019）。判别器通过区分真实时间序列数据与生成数据来进行训练，其目标是产生与真实数据无法区分的合成数据。TimeGAN（Yoon等，2019）作为一种基于生成对抗网络的架构，被提出用于生成逼真的时间序列数据，其结合了无监督模型的灵活性与有监督模型的可控性。该架构利用嵌入函数和恢复函数从时间序列数据中提取高维特征，随后将这些特征输入序列生成器和序列判别器进行对抗训练。另一项研究提出了一种基于循环神经网络（RNN）的生成对抗网络架构，用于生成实值多维时间序列（Takahashi等，2019）。该研究引入了两种变体：递归生成对抗网络（RGAN）和递归条件生成对抗网络（RCGAN），其中RGAN生成实值数据序列，而RCGAN则基于特定输入条件生成序列。RGAN与RCGAN的判别器和生成器均基于简单的循环神经网络架构实现

sequences, and RCGAN generates sequences conditioned on specific inputs. The discriminators and generators of both RGAN and RCGAN are based on simple RNN architectures.

此外，有研究提出了一种基于深度神经网络的方法来建模金融时间序列数据（Takahashi等，2019）。该方法通过数据驱动的方式学习数据特性并生成具有真实性的数据，同时保留了金融时间序列的统计特征，包括非线性可预测性、厚尾收益分布、波动率聚集效应、杠杆效应、从粗到细的波动相关性以及非对称收益/损失模式等特征

Furthermore, a deep neural network-based approach was proposed for modeling financial time series data (Takahashi et al., 2019). This approach learns the properties of the data and generates realistic data in a data-driven manner, while preserving statistical characteristics of financial time series such as nonlinear predictability, heavy-tailed return distributions, volatility clustering, leverage effect, coarse-to-fine volatility correlations, and asymmetric return/loss patterns.

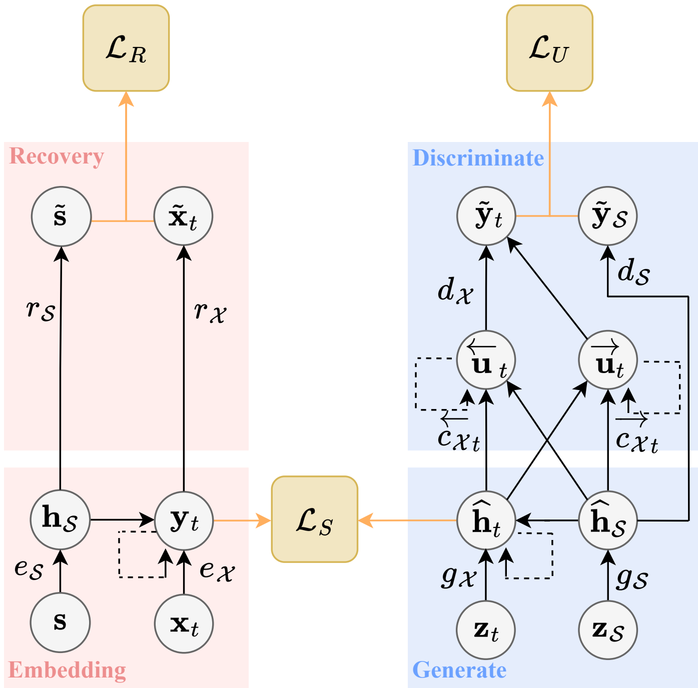


Fig. 6 The overview of TimeGAN model

这些研究凸显了生成对抗网络在时间序列预测领域的应用，特别是在生成合成时间序列数据及捕捉现实世界时间序列数据特征方面的作用

These studies highlight the application of GANs in TSF, specifically in generating synthetic time series data and capturing the characteristics of real-world time series data.

3.2 Model without Explicit Structure

3.2.1 Integrated Model

As widely known, recurrent neural networks (RNNs) are often considered suitable for sequence modeling, and the chapter on sequence modeling in classic deep learning textbooks is titled “Sequence Modeling: Recurrent and Recursive Nets” (Heaton, 2018).

众所周知，循环神经网络（RNNs）通常被认为适用于序列建模任务，经典深度学习教材中关于序列建模的章节便以《序列建模：循环与递归网络》为标题（Heaton, 2018）。时间序列数据天然属于序列建模范畴，因此RNN、LSTM、GRU等模型理应适用于解决时间序列相关任务。然而在音频合成、词级语言建模和机器翻译等领域，卷积架构却取得了最先进的精度表现（Ba等, 2018a），这一现象引发广泛关注，促使研究者探索如何将卷积架构应用于序列建模领域。集成模型由此应运而生

Time series naturally falls within the realm of sequence modeling tasks, and therefore, RNNs, LSTM, GRU, and similar models are expected to be applicable to solve time series-related tasks. However, convolutional architectures have achieved state-of-the-art accuracy in tasks such as audio synthesis, word-level language modeling, and machine translation (Bai et al., 2018a), which has garnered significant attention and led to inquiries on how to apply convolutional architectures in the domain of sequences. Integrated models have emerged as a solution.

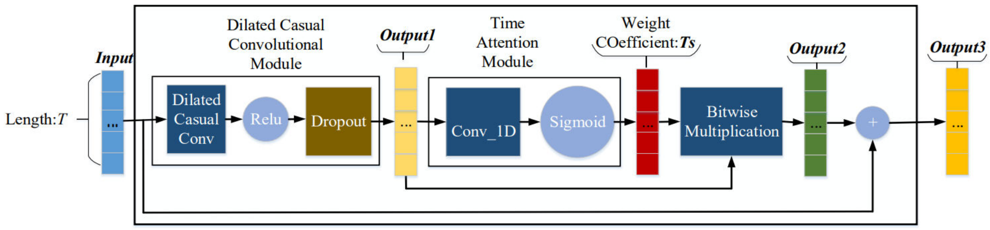


Fig. 7 The overview of TATCN model

Integrated models can combine the strengths of individual model architectures, with each focusing on learning features it excels at, resulting in improved performance. For example, convolutional architectures excel at learning local feature patterns, while recurrent architectures excel at learning temporal dependencies between nodes. Integrated models have also found various applications in time series tasks (Bai et al., 2018a; Shi et al., 2015; Asiful et al., 2018). In (Shi et al., 2015), precipitation forecasting was modeled as a spatio-temporal sequence prediction problem, where a convolutional architecture was designed to replace fully connected layers in LSTM for sequence modeling, effectively leveraging the advantages of both convolutional and recurrent architectures. Similarly, Asiful et al. (2018) integrated multiple network architectures, namely LSTM and GRU, for stock prediction. In this model, the input was first fed into the LSTM layer, then into the GRU layer, and finally into a dense network.

集成模型能够融合不同模型架构的优势，使每种架构专注于学习其擅长的特征模式，从而提升整体性能。例如，卷积架构擅长捕捉局部特征模式，而循环架构则擅长学习节点间的时间依赖性。此类集成模型在时间序列任务中已展现出广泛的应用价值（Bai等，2018a；Shi等，2015；Asiful等，2018）。在Shi等（2015）的研究中，降水预测被建模为时空序列预测问题，研究者设计了一种卷积架构替代LSTM中的全连接层进行序列建模，有效结合了卷积与循环架构的双重优势。类似地，Asiful等（2018）通过整合LSTM和GRU两种网络架构构建了股票预测模型，该模型先将输入数据传递至LSTM层，再经GRU层处理，最终通过密集网络输出预测结

3.2.2 Cascade Model

Cascade networks, which are widely used in deep neural networks, especially in Computer Vision (CV) domain (Cai and Vasconcelos, 2018), have multiple applications. A cascade network typically consists of multiple components, each serving a different function, collectively forming a deeper and more powerful network model. The components in a cascade model can be either identical or different. When the components are different, each component has a specific role and function. If the components are the same, it means that a particular module or the entire network is repeated several times. When the same component is repeated multiple times, its concept is somewhat similar to the iterative approach used in solving optimization problems.

级联网络在深度神经网络中具有广泛应用，尤其在计算机视觉领域（Cai和Vasconcelos，2018）。这类网络通常由多个组件构成，每个组件承担不同功能，共同形成更深层、更强大的网络模型。级联模型中的组件既可以是同构的也可以是异构的。当组件异构时，每个组件都具有特定的角色和功能；若组件同构，则意味着特定模块或整个网络被多次重复。当相同组件被多次重复时，其概念在某种程度上类似于优化问题求解中采用的迭代方法

In the field of TSF, there are not many works specifically known for their cascade models. However, the concept of cascade is widely applied in various network model architectures. Firstly, stacking multiple identical modules or the entire network can be

在时间序列预测（TSF）领域，专门以其级联模型设计而闻名的研究工作并不多见。然而，级联的概念在各种网络模型架构中得到了广泛应用。首先，堆叠多个相同模块或整个网络即可视为运用了级联思想，这在 Transformer 系列模型中有充分体现。此外，一些模型采用了特殊设计的级联方法，旨在确保信息以特定方式流动，从而实现独特的效果

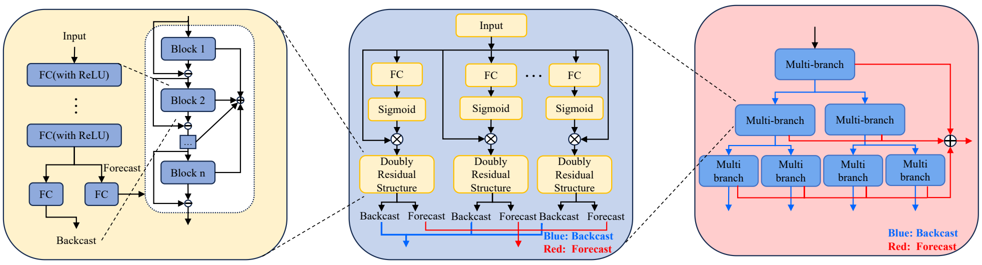


Fig. 8 The overview of TreeDRNet model

considered as utilizing the cascade idea, as seen in the Transformer series (Makridakis, 1978; Cleveland et al., 1990; Li et al., 2019a; De Livera et al., 2011). Additionally, some models (Zhou et al., 2022c) incorporate specially designed cascade approaches to ensure the flow of information in a specific manner, thereby achieving unique effects.

4 Series Components and Enhanced Feature Extraction Methodology

在前文章节中，我们已系统阐述了构建深度时序特征模型的五大主流范式，为研究者提供了一条理解与构建深度学习模型的清晰路径。然而，仅从宏观层面理解与构建深度时序特征模型仍显不足。本章将深入探讨时序特征学习的方法论体系，通过预训练、分解、提取与精炼的递进过程，使模型能够更精准地捕捉数据的本质表征，这一过程严格遵循数据的内在特性进行设计

In the previous sections, we have provided a comprehensive overview of five prominent paradigms for constructing DTSF models. These paradigms offer researchers a concise pathway to understanding and building DL models. However, a macroscopic understanding and construction of DTSF models alone is insufficient. This chapter delves into the methodological aspects of learning temporal features, which enable models to better capture the underlying representations of the data, emphasizing a pre-training, decomposition, extraction, and refinement process that aligns closely with the intrinsic nature of data.

本章分为两部分。首先剖析现实世界中时间序列数据的构成要素，随后深入探讨四种理论基础扎实、领域内表现突出的成熟特征提取方法。这些方法有助于更全面地理解时间序列数据及其核心特征

The chapter is divided into two parts. It begins by dissecting the constituents of time series data in the real world. Subsequently, it proceeds to provide an in-depth exploration of four well-established feature extraction methods with strong theoretical foundations and notable performance in the field. These methods facilitate a richer understanding of time series data and its essential features.

4.1 Components of a Time Series

一般而言，时间序列数据可分解为三个主要组成部分：趋势项、季节性成分以及残差或白噪声（Shumway等, 2000），如图9所示

In general, time series data can be decomposed into three main components: trend, seasonality, and residuals or white noise (Shumway et al., 2000), as illustrated in Figure 9.

4.1.1 Trend

它代表了时间序列数据中的长期变化，反映了数据在较长时间跨度内的整体增长或减少趋势 (Montgomery et al., 2015)。例如，人口数量的逐年增长呈现出一种上升趋势 (Adhikari and Agrawal, 2013)，而多个风季中风电发电量的持续增长也可被视为一种上升趋势

Represents the long-term changes in the time series data and reflects the overall growth or decline of the data over an extended period (Montgomery et al., 2015). For example, the increase in population over the years exhibits an upward trend (Adhikari and

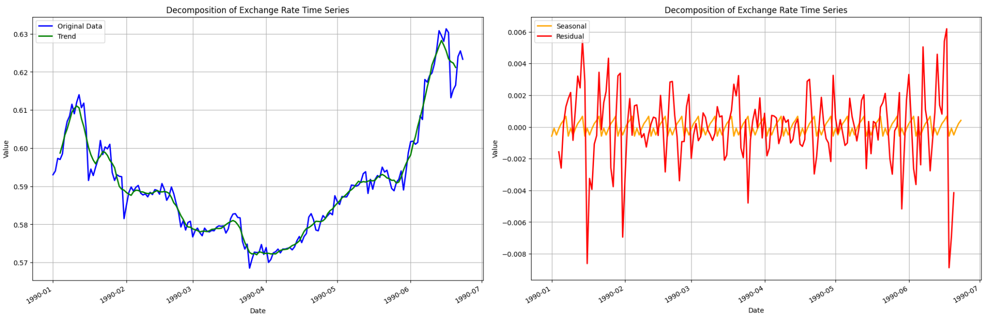


Fig. 9 Components of the time series. The data is sourced from the Exchange-Rate dataset spanning from January 1, 1990, to June 23, 1990. The blue line represents the original data, the green indicates the trend, the yellow represents seasonality, and the red signifies the residuals

时间序列的构成要素。数据源自1990年1月1日至1990年6月23日的汇率数据集。蓝色线条代表原始数据，绿色表示趋势项，黄色表征季节性成分，红色则对应残差项

Agrawal, 2013), and the growing wind power generation during multiple windy seasons can also be considered an upward trend.

4.1.2 Seasonality

Refers to the periodic variations observed in time series data, often caused by seasonal, monthly, weekly, or other time unit influences. For instance, the number of tourists and ice cream sales tend to increase during long vacations or in the summer.

指在时间序列数据中观察到的周期性波动，通常由季节性、月度、周度或其他时间单位的影响所导致。例如，长假期间或夏季，游客数量与冰淇淋销量往往呈现上升趋势

4.1.3 Residuals

Represent the part of the data that cannot be explained by the trend and seasonality components (Maronna et al., 2019). They capture the random fluctuations or noise remaining after the decomposition of trend and seasonality. Residuals reflect the short-term fluctuations and irregularities that have not been modeled in the time series data. Additionally, residuals exhibit some autocorrelation, which can help us identify and adjust for potential flaws in the model, further enhancing the quality and reliability of forecasting.

表示数据中无法被趋势和季节性成分解释的部分（Maronna等人，2019年）。它们捕捉了在分解趋势和季节性后剩余的随机波动或噪声。残差反映了时间序列数据中尚未建模的短期波动和不规则性。此外，残差表现出一定的自相关性，这有助于我们识别并调整模型中潜在的缺陷，从而进一步提升预测的质量和可靠性

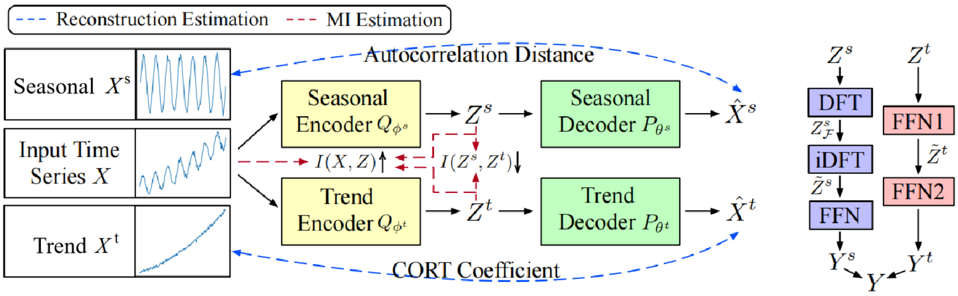
In the real world, time series data contains discrete information and is non-stationary, meaning that its mean and variance are not constant over time. By decomposing the data into its constituent parts, we gain a better understanding of the data's structure, identify long-term trends and periodic variations, and distinguish them from random noise. These decomposition components aid in making more accurate forecasts, uncovering hidden patterns, extracting useful information, and providing insights into the mechanisms and regularities underlying the time series data.

在现实世界中，时间序列数据包含离散信息且具有非平稳性，即其均值与方差随时间推移并非恒定不变。通过将数据分解为构成成分，我们能更深入地理解数据结构，识别长期趋势与周期性波动，并将其与随机噪声区分开来。这些分解成分有助于提升预测精度、揭示潜在模式、提取有效信息，并为理解时间序列数据的内在机制与规律提供洞见

4.2 Methodology for Enhanced Feature Extraction

Numerous studies have been dedicated to improving the model architecture and refining its components in DTSF. These studies aim to enhance the predictive performance

诸多研究致力于在深度时间序列预测(Deep Time Series Forecasting, DTSF)领域中改进模型架构并完善其组件。这些研究旨在通过优化或替换用于特征提取与特征学习的方法来提升模型的预测性能。为了实现精确预测，深入习得时间序列的表示特征至关重要，并且充足的信息对于训练高质量的模型参数必不可少



(a) Seasonal-trend representations learning (b) Predictor

Fig. 10 The overview of LaST model

of models by optimizing or replacing the methods used for extraction and feature learning. To achieve accurate predictions, it is crucial to learn time series representation features thoroughly, and sufficient information is essential for training high-quality model parameters.

近年来，关于深度时序预测（DTSF）的研究成果在数据处理与分量建模方面展现出显著变革。其中，将时间序列分解为核心组分进行解析已成为主要研究方向，这为深入探究趋势特征与季节维度提供了更全面的途径。值得注意的是，将时域数据转换至频域的处理方法已被证实能更有效地实现特征区分。此外，探索非端到端的技术路径，并设计适配的数据预训练方案以解决目标任务与数据间的潜在失配问题，同样具有重要研究价值。下文将系统阐述DTSF中增强特征提取与学习能力的主要方法体

In recent years, influential works on DTSF have shown significant changes in data processing and component modeling. Notably, decomposing time series into its major components for analysis has been a primary focus, facilitating a more comprehensive exploration of trends and seasonal dimensions. Furthermore, transforming time-domain data into the frequency domain has proven to be more effective in feature differentiation. Additionally, exploring non-end-to-end approaches and devising suitable data pre-training methods to address the potential mismatch between the target task and the data is also a valuable consideration. In the following sections, we will introduce the primary methodologies for enhancing feature extraction and learning in DTSF.

4.2.1 Dimension Decomposition

维度分解在时间序列预测领域具有重要作用。该方法通过将数据分解为趋势项、季节项和残差等构成维度或分量来实现分析

Dimension decomposition plays a vital role in the realm of TSF. It involves breaking down the data into its constituent dimensions or components, such as trends, seasonal patterns, and residuals.

在当前研究中，部分工作将编码器-解码器架构与季节性-趋势性分解技术相结合 (Wu et al., 2021; Zhou et al., 2022b; Zhang et al., 2022b; Wang et al., 2022; Zhu et al., 2023; Tang and Zhang, 2023; Cao et al., 2023b; Peng et al., 2023)。Wu 等人 (2021) 设计了一种内部分解模块，旨在赋予深度预测模型内在的渐进式分解能力。随后，Zhou 等人 (2022b) 在 FEDformer 框架中提出了一种基于季节性-趋势性的频率增强分解型 Transformer 架构。此外，Wang 等人 (2022) 提出了 LaTS 模型，利用变分推理挖掘潜在空间中的季节性与趋势性特征。Zhang 等人 (2022b) 则提出了 TDformer 模型，该模型使用 MLP（多层感知机）建模趋势项，并采用傅里叶注意力机制来模拟季节性成分。尤为值得注意的是，Zhu 等人 (2023) 设计了一种将输入序列分解为多尺度趋势分量和残差分量的方法，并将学习到的这些分量特征求和作为模型输出。在近期工作中，通过采用对比学习和一种增强型分解架构，成功克服了捕捉外窗（outer-window）变化这一挑战 (Park et al., 2024)。据观察，分解网络能够显著助力于基于对比损失（contrastive loss）的长时表示学习，从而提升长期预测的性能

In current research, some works have integrated encoder-decoder architectures with seasonal-trend decomposition (Wu et al., 2021; Zhou et al., 2022b; Zhang et al., 2022b; Wang et al., 2022; Zhu et al., 2023; Tang and Zhang, 2023; Cao et al., 2023b; Peng et al., 2023). Wu et al. (2021) in the similar work, devised an internal decomposition block to endow deep forecasting model with intrinsic progressive decomposition capability. Subsequently, Zhou et al. (2022b) proposed a seasonal-trend-based frequency enhanced decomposition Transformer architecture in the FEDformer framework. Additionally, Wang et al. (2022) introduced the LaTS model, leveraging variational inference to unravel latent space seasonal trend features, and Zhang et al. (2022b) presented the TDformer model, using MLP to model trends and Fourier attention to simulating seasonality. Notably, Zhu et al. (2023) designed an approach to decompose input sequences into trend and residual components across multiple scales,

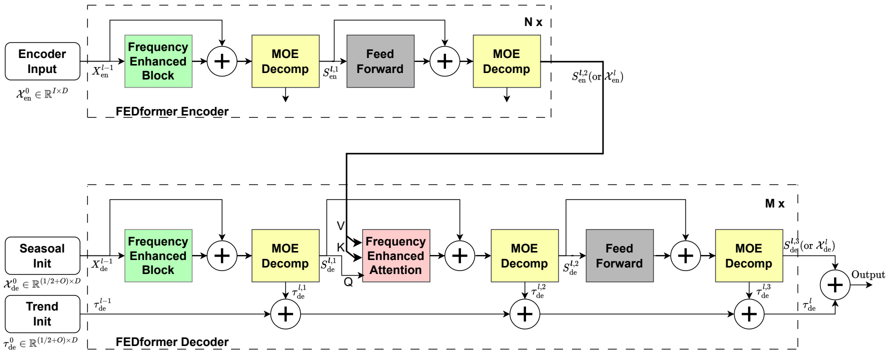


Fig. 11 The overview of FEDformer model

which summed the learned features as the model output. In recent work, the challenge of capturing outer-window variations was overcome by employing contrastive learning and an enhanced decomposition architecture (Park et al., 2024). It is observed that decomposition networks can significantly benefit contrastive loss learning of long-term representations, thereby enhancing the performance of long-term forecasting.

The significance of dimension decomposition lies in its ability to delve into and capture the inherent components or dimensions within time series data. On one hand, it aids in isolating and extracting latent patterns in time series data for identification and analysis. On the other hand, it isolates individual features that influence the overall behavior, allowing for a more focused analysis of each constituent part. This contributes to understanding the impact of each feature on the overall time series. Furthermore, decomposing data dimensions enhances the interpretability of TSF models, which facilitates a better understanding of the influence of different components on overall temporal behavior. As a relatively universal method in time series analysis, dimension decomposition plays a foundational yet crucial role in enhancing feature extraction methodologies.

维度分解的意义在于其能够深入挖掘并捕捉时间序列数据中固有的成分或维度。一方面，它有助于分离和提取时间序列数据中的潜在模式以进行识别与分析；另一方面，该方法能将影响整体行为的独立特征进行隔离，从而实现对各个组成部分的更聚焦分析。这有助于理解每个特征对整体时间序列的影响机制。此外，数据维度的分解增强了时序预测模型的可解释性，有利于更清晰地理解不同成分对整体时序行为的作用机理。作为时间序列分析中相对普适的方法，维度分解在优化特征提取方法论方面发挥着基础而关键的作用

4.2.2 Time-Frequency Conversion

The time-frequency domain conversion plays a crucial role in deep learning-based time series forecasting tasks. It refers to converting the time-domain data into its frequency-domain representation, enabling a more effective analysis of the frequency, spectral characteristics, and dynamic variations within time series data.

时频域转换在基于深度学习的时间序列预测任务中具有关键作用。该技术通过将时域数据转化为频域表示，能够更有效地分析时间序列数据中的频率成分、频谱特征及动态变化规律

In current research, the time-frequency domain conversion finds extensive application in the preprocessing and feature extraction of time series data (Kourentzes et al., 2014; Chen et al., 2023; Sun and Boning, 2022). This method reveals the components of the data at different frequencies and aids in identifying repetitive patterns, periodic trends, and frequency-domain features such as seasonal patterns or periodic oscillations (Zhou et al., 2022b). Converting time series data into spectrograms provides an overview of the data's distribution in the frequency domain, facilitating the identification of major frequency components and the shape of the spectrum. This is particularly valuable for capturing the overall spectral characteristics of signals and

在当前研究中，时频域转换技术被广泛应用于时间序列数据的预处理和特征提取环节 (Kourentzes et al., 2014; Chen et al., 2023; Sun and Boning, 2022)。该方法能揭示数据在不同频率下的分量，并有助于识别重复模式、周期性趋势以及诸如季节性模式或周期性振荡等频域特征 (Zhou et al., 2022b)。将时间序列数据转换为频谱图（spectrograms），可概览数据在频域的分布情况，从而便于识别主要频率分量和频谱形状。这对于捕捉信号的整体频谱特性及其在频率维度上的主要波动模式尤为有价值。Cao 等人 (2020) 在其工作中采用了 StemGNN，旨在光谱域中联合捕获多变量时间序列预测所需的序列间相关性和时序依赖性。在近期工作中，Yi 等人 (2023) 提出了一种基于频域多层感知机（Frequency-Domain MLP）的、简单而高效的时间序列预测架构，命名为 FreTS。该架构主要由两个阶段组成：域转换（domain conversion）和频率学习（frequency learning），这两个阶段协同作用，增强了在跨序列（inter-series）与序列内（intra-series）两尺度下对通道相关性和时序相关性的学习能力

the primary fluctuation patterns across frequencies. In their work, (Cao et al., 2020) employ StemGNN to jointly capture inter-sequence correlations and temporal dependencies in the spectral domain for multivariate time series forecasting. In recent work, Yi et al. (2023) proposed a simple yet effective time series forecasting architecture, named FreTS, based on Frequency-Domain MLP. It primarily consists of two stages, domain conversion and frequency learning, which enhance the learning of channel and temporal correlations across both inter-series and intra-series scales.

此外，采用时频域转换技术有助于降低噪声与干扰的影响（Zhou等，2022a；Gu等，2021）。在特定时间序列预测场景中，噪声可能污染数据，导致模型预测性能下降。Zhou等（2022a）在FiLM模型中创新性地引入了频率增强层来解决这一问题，通过结合傅里叶分析与低秩矩阵近似实现噪声抑制，有效降低了噪声信号的干扰并缓解了过拟合现象。研究表明，将时域数据转换至频域后实施滤波、去噪等频域操作，对于削弱噪声影响具有显著效果

Furthermore, employing time-frequency domain conversion can help reduce the impact of noise and interference (Zhou et al., 2022a; Gu et al., 2021). In specific time series forecasting scenarios, noise may affect the data, resulting in a decline in the model's predictive performance. In the FiLM model, Zhou et al. (2022a) introduced a Frequency Enhancement Layer to address this issue. They achieved noise reduction by combining Fourier analysis and low-rank matrix approximation, which minimized the influence of noise signals and mitigated overfitting problems. Apparently, converting time-domain data into the frequency-domain, along with operations like filtering and denoising in the frequency domain, proves effective in lessening the impact of noise.

时频域转换的重要性在于为数据分析提供了一种全面而细致的方法，该方法能够揭示时间序列中隐藏的频率特征与动态变化规律。该技术在时间序列预测领域得到了广泛应用，是提升预测性能和理解时序数据复杂性的关键方法论

The importance of time-frequency domain conversion lies in providing a comprehensive and detailed approach to data analysis, which is capable of unveiling the hidden frequency characteristics and dynamic changes within time series. This technique has been widely employed in the domain of TSF, representing a crucial methodology for enhancing predictive performance and comprehending the intricacies of time series data.

4.2.3 Pre-training

与自然语言相比，时序数据的信息密度较低，需要更长的序列才能捕捉其时序模式。此外，时序数据还存在时序动态性、快速演化以及长短期效应并存等挑战。由于预训练与目标领域间可能存在不匹配，下游任务性能可能受到影响。近期时间序列预测研究在自监督与无监督预训练方面进行了创新尝试，取得了显著成果（Rebjock等，2021；Sun等，2021；Sarkar与Etemad，2020；Cheng等，2020）。在某些应用场景中，可考虑采用采样预训练方法加以解

Compared to natural language, temporal data exhibits lower information density, necessitating longer sequences to capture temporal patterns. Additionally, temporal data also exist challenges such as temporal dynamics, rapid evolution, and the presence of both long and short-term effects. Due to potential mismatches between pre-training and target domains, downstream performance might suffer. Recent endeavors in TSF involve novel attempts at self-supervised and unsupervised pre-training, yielding promising results (Rebjock et al., 2021; Sun et al., 2021; Sarkar and Etemad, 2020; Cheng et al., 2020). In certain scenarios, the adoption of sampling pre-training methods could be considered.

对比预训练。由于预训练与目标领域间可能存在不匹配，时间序列预训练面临一项独特挑战，即可能导致下游任务性能下降。虽然领域自适应方法能够缓解这些变化（Berthelot等人，2022；Singh，2021），但大多数方法对于预训练而言仍被视为次优解，因其通常需要目标领域的直接样本。为解决这一问题，这些方法需要在预训练期间不依赖任何目标领域样本的情况下，自适应地调整以适应目标领域多样化的时序动态特征

Contrastive pre-training. Due to potential mismatches between pre-training and the target domain, there is a unique challenge in time series pre-training that may lead to diminished downstream performance. While domain adaptation methods can alleviate these changes (Berthelot et al., 2022; Singh, 2021), most approaches are considered suboptimal for pre-training as they often require direct examples from the target domain. To address this, these methods need to adapt to the diverse temporal dynamics of the target domain without relying on any target examples during pre-training.

Contrastive learning, a form of self-supervised learning, aims to train an input encoder to map positive sample pairs closer and negative pairs apart (Oord et al., 2018). In time series, if the representations based on time and frequency for the same

对比学习（Contrastive Learning）作为自监督学习的一种形式，其核心目标是训练一个输入编码器（encoder），将正样本对在表示空间中拉近，同时将负样本对推远（Oord et al., 2018）。在时间序列领域，如果同一实例（instance）基于时域和基于频域的表征在时频空间（time-frequency space）中彼此接近，则表明其底层特征或属性存在某种相似性或一致性（Zhang et al., 2022a）。Zhang 等人 (2022a) 提出了在预训练中追求时频一致性（Time-Frequency Consistency, TF-C） 的需求，这要求将一个样本基于时间的邻域（time-based neighborhood）嵌入到与其基于频率的邻域（frequency-based neighborhood）相近的位置。该研究采用基于频率的对比增强（frequency-based contrastive enhancement）技术，旨在利用丰富的光谱信息（spectral information）并探索时间序列中的时频一致性。对比式预训练能够为预测任务提供鲁棒的特征表示，从而有助于提升模型性能（enhanced model performance） 和泛化能力（generalization）

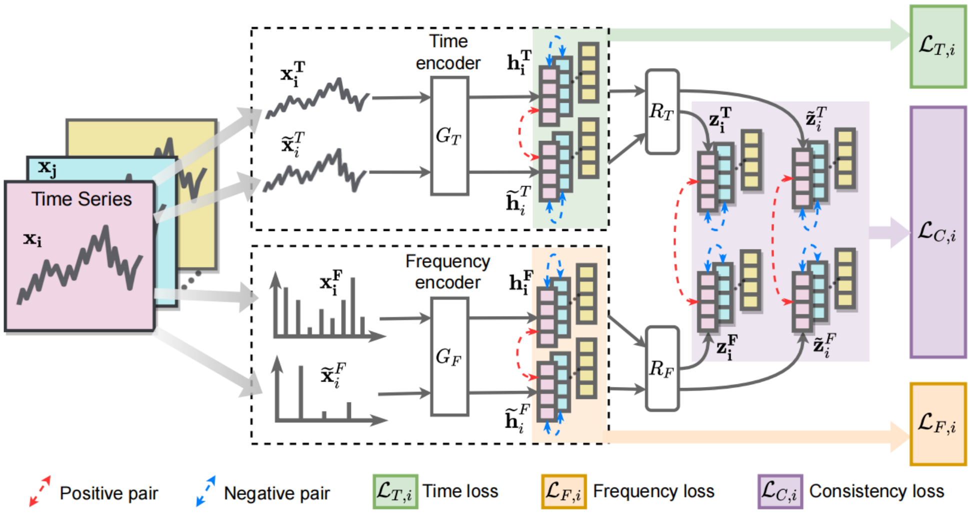


Fig. 12 The overview of TF-C

instance are close in the time-frequency space, it suggests a certain similarity or consistency in their features or attributes. Zhang et al. (2022a). proposed the need for Time-Frequency Consistency (TF-C) in pre-training, which involves embedding the time-based neighborhood of an example close to its frequency-based neighborhood. This work employs frequency-based contrastive enhancement to leverage rich spectral information and explore time-frequency consistency in time series. Contrastive pre-training can provide robust feature representations for forecasting tasks, contributing to enhanced model performance and generalization.

Masking Pre-training. Time series data is often continuous, ordered, but practically exhibits incompleteness. Additionally, real-world time series data commonly contains noise and uncertainty, necessitating models to possess robustness in dealing with such uncertainties. To address these crucial challenges in practice, the masking mechanism is regarded in some studies as an effective approach to enhance feature extraction.

掩码预训练。时间序列数据通常具有连续性、有序性，但在实际应用中往往呈现不完整性。此外，现实世界的时间序列数据普遍存在噪声和不确定性，这就要求模型必须具备处理此类不确定性的鲁棒性。为应对这些实践中的关键挑战，部分研究将掩码机制视为增强特征提取能力的有效途径

In the work STEP, Shao et al. (2022) designed an unsupervised pre-training model for time series based on Transformer blocks. The model employs a masked autoencoding strategy for training, which effectively learns temporal patterns and generates segment-level representations. These representations provide contextual information for subsequent inputs, facilitating the modeling of dependencies between short-term time series. The Ti-MAE model (Li et al., 2023b) exhibits analogous efficacy in this regard. In the pre-training model SimMTM, Dong et al. (2023) highlighted that randomly masking parts of the data severely disrupts temporal variations. They relate masking modeling to manifold learning and propose a Simple pre-training framework for Masked Time-series Modeling.

在STEP研究中，Shao等人（2022）设计了一种基于Transformer模块的无监督时间序列预训练模型。该模型采用掩码自编码策略进行训练，能有效学习时序模式并生成片段级表征。这些表征为后续输入提供上下文信息，有助于建立短期时间序列间的依赖关系模型。Ti-MAE模型（Li等，2023b）在此方面展现出相似效能。在预训练模型SimMTM中，Dong等人（2023）指出随机掩蔽部分数据会严重破坏时序变化特征。他们将掩蔽建模与流形学习相关联，提出了一个面向掩蔽时间序列建模的简易预训练

In summary, Masking pre-training simulates incompleteness and noise by masking some data points, enabling the model to learn how to handle partially missing information during the pretraining phase. This methodology can enhance the model's

总而言之，掩码预训练（Masking pre-training）通过遮蔽部分数据点来模拟数据不完整性（incompleteness）和噪声（noise），从而使模型在预训练阶段学会如何处理部分缺失的信息。这种方法能够增强模型捕获长期依赖关系的能力、提高其对数据不确定性的容忍度（tolerance），并改善其整体泛化性能（generalization performance）

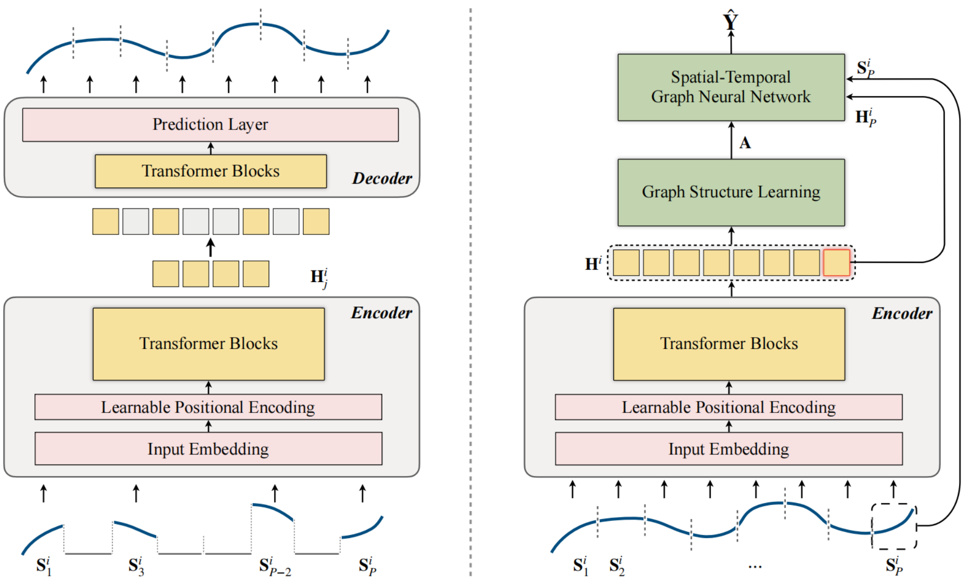


Fig. 13 The overview of STEP

ability to capture long-term dependencies, increase tolerance to data uncertainty, and improve overall generalization performance.

4.2.4 Patch-based segmentation

在近期的深度时间序列预测（DTSF）研究中，尤其是基于Transformer模型的工作中，基于分块（patch-based）的数据组织方式已得到广泛应用（Nie等，2022；Lin等，2023；Zhang等，2023b；Ekambaram等，2023；Xue等，2023；Gong等，2023）。 采用分块策略有助于增强模型的局部感知能力。通过将长时间序列分割为较小的数据块，模型能够更有效地捕捉序列中的短期与局部模式，从而提升其对序列复杂动态特征的理解。 与此同时，多元变量间的关联性可产生信息增益。主要挑战在于如何学习各变量间的相互关系，并将有效信息引入模型，同时避免冗余信息对模型训练过程造成干扰

In recent DTSF works, especially those of the Transformer models, the adoption of patch-based data organization has become prevalent (Nie et al., 2022; Lin et al., 2023; Zhang et al., 2023b; Ekambaram et al., 2023; Xue et al., 2023; Gong et al., 2023). It is advantageous to enhance the model's local perception capabilities by employing a patch-based strategy. Through segmenting long time series into smaller patches, the model becomes more adept at capturing short-term and local patterns within the sequence, thereby augmenting its comprehension of complex dynamics in the sequence. Simultaneously, the relationships among multivariate variables can yield information gain. Challenges lie primarily in how to learn the relationships among individual variables and introduce valid information into the model, while avoiding redundant information that may interfere with the model training process.

聂等人（2022）提出了PatchTST模型，该模型将时间序列分割为子序列级别的片段（patch），作为Transformer的输入标记。他们采用通道独立的方式对每个变量进行单独建模，这种方法不仅能在嵌入层有效保留各变量的局部语义信息，还能关注更长期的历史特征。此外，借助通道独立的特性，通过图建模方法（张等人，2023c）可以进一步学习单变量间潜在的特征关联，实现图中全局标记表征的空间聚合

Nie et al. (2022) proposed the PatchTST model, where they segment time series into subseries-level patches, serving as input tokens for the Transformer. They independently model each channel to represent a single variable. This channel-independent approach not only effectively preserves local semantic information for each variable in the embedding but also focuses on a more extended history. Furthermore, leveraging the channel-independent characteristics, potential feature correlations between single variables can be further learned through graph modeling methods (Zhang et al., 2023c). It allows for spatial aggregation of representations for global tokens in the graph.

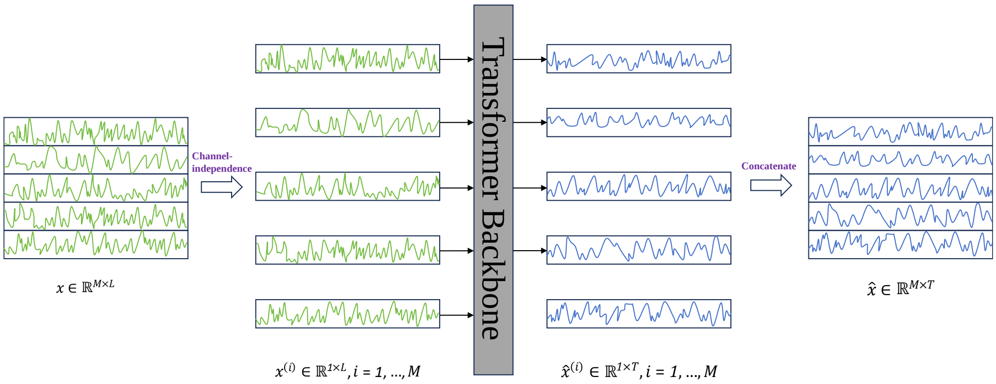


Fig. 14 The overview of PatchTST

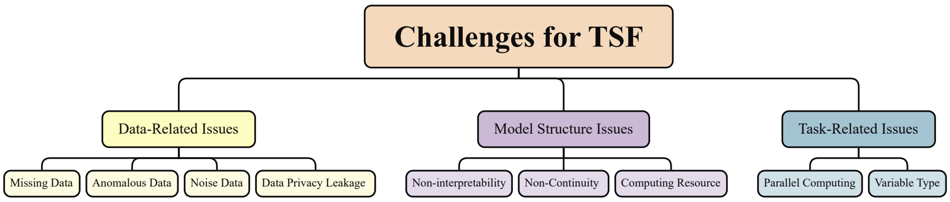


Fig. 15 Challenges in Time Series Forecasting

While the modeling emphasis varies across different works, there is a common consideration of employing methods that utilize subseries-level patches to process the raw time series data. This approach proves highly beneficial for capturing and learning the local features of the data. The patch-based segmentation method introduces another methodology for TSF. Additionally, channel independence emerges as a viable avenue for exploring multivariate time series forecasting.

尽管不同研究在建模侧重点上存在差异，但普遍考虑采用基于子序列级片段的方法来处理原始时间序列数据。这种方法被证明对捕捉和学习数据的局部特征极为有效。基于片段的分割策略为时间序列预测引入了另一种方法论。此外，通道独立性成为探索多元时间序列预测的一条可行路径

5 Challenges and Prospects

We have investigated the neural network architectures, feature extraction and learning approaches, and significant experimental datasets of deep learning models in the context of TSF. While DTSF models have demonstrated remarkable achievements across diverse domains in recent years, certain challenging issues remain to be addressed, which point towards potential future research directions. We summarize these challenges and propose viable avenues as follows. We classify the challenges into three main categories: data features, model structure, and task-related issues. Within each category, we highlight several representative challenges. Figure 15 illustrates an overview of these challenges.

我们研究了时间序列预测（TSF）背景下深度学习模型的神经网络架构、特征提取与学习方法，以及重要的实验数据集。尽管深度时间序列预测（DTSF）模型近年来在多个领域取得了显著成果，但仍存在若干亟待解决的挑战性问题，这些难题为未来研究指明了潜在方向。现将这些挑战归纳总结并提出可行的解决路径如下。我们将挑战分为三大类：数据特征、模型结构和任务相关问题。在每个类别中，我们重点阐述了若干具有代表性的挑战。图15展示了这些挑战的总体概览

Table 3 Time Series Datasets in Primary Domains. The table summarizes commonly used datasets and indicates whether they are multivariate, which implies temporal alignment with known timestamps

Domain	Datasets	Variants	Data Time Range	Data Granularity	Multi/Uni	Authors
Energy	ETTh1	7	2016 - 2018	1h	Multi+Uni	Zhou et al.
	ETTm1	7	2016 - 2018	15m	Multi+Uni	Zhou et al.
	Electricity	321	2011 - 2014	1h	Multi+Uni	-
	Wind	28	1986 - 2015	1h	Uni	-
	Solar-Energy	137	2006 - 2006	10m	Multi+Uni	Solar
Healthcare	ILI	7	2002 - 2021	1w	Uni	-
	MIT-BIH	2	1975 - 1979	360Hz	Uni	George
Transportation	Traffic	862	2015 - 2016	1h	Uni	Caltrans
	PeMSD4	307	2018/1	5m	Multi	Chen et al.
	PeMSD7	228	2012/5			
	PeMSD8	170	2016/7			
Meteorology	Weather1	12	1981 - 2010	1h	Uni	-
	Weather2	21	2020 - 2021	10m	Multi+Uni	Sparks et al.
	Temperature Rain	2	2015 - 2017	1d	Multi+Uni	Rakshitha et al.
Economics	Exchange-Rate	8	1990 - 2016	1d	Uni	Lai et al.
	LOB-ITCH	149	2010 - 2010	1ms-10min	Uni	Adamantios et al.
	Dominick	25	1989 - 1994	1w	Uni	Godahewa et al.

5.1 Challenges

5.1.1 Lack of Data Privacy Protection and Completeness

联邦学习（Federated Learning，FL）在时间序列预测（TSF）领域正日益受到重视，其核心价值在于解决本地数据体量庞大与信息交换中的隐私保护难题（McMahan等，2017）。该技术允许多个参与方在不共享敏感原始数据的前提下协同训练模型。在TSF任务中，各参与方可利用本地时间序列数据进行模型训练，通过联邦学习算法聚合局部模型参数，最终获得全局预测模型。这种分布式学习机制有效保障了数据隐私，规避了集中式数据存储与传输可能引发的隐私泄露风险。 当前研究主要集中在负荷检测（Gao等，2021；Ta¨ık与Cherkaoui，2020；Briggs等，2022）、交通速度与流量（Zhang等，2021；Liu等，2020）、能源消耗（Zhang等，2020；Savi与Olivadese，2021）以及通信网络（Subramanya与Riggio，2021；D´ıaz Gonz´alez，2019）等领域，而在其他领域探索可行解决方案仍是一个待开拓的方向。值得注意的是，联邦学习通过整合分布式数据源的多样性，能够显著提升模型的泛化能力与预测精度。因此，该技术为大规模、安全高效的时间序列预测与分析提供了极具前景的解决方案

Federated learning (FL) is gaining momentum in the field of TSF, primarily addressing challenges associated with large local data volumes and privacy concerns during information exchange. With FL, multiple participants can collaboratively train models without the need to share sensitive raw data (McMahan et al., 2017). In TSF tasks, each participant can leverage their local time series data for model training. Through FL algorithms, the parameters of local models are aggregated to obtain a global predictive model. This distributed learning process ensures privacy protection, mitigating the risks of privacy breaches associated with centralized data storage and transmission. Current research efforts predominantly focus on load detection (Gao et al., 2021; Taïk and Cherkaoui, 2020; Briggs et al., 2022), traffic speed and flow (Zhang et al., 2021; Liu et al., 2020), energy consumption (Zhang et al., 2020; Savi and Olivadese, 2021), and communication networks (Subramanya and Riggio, 2021; Díaz González, 2019), among others. Exploring feasible solutions in other domains remains an open avenue. Furthermore, federated learning harnesses the diversity of distributed data sources, thereby enhancing model generalization and prediction accuracy. Hence, federated learning holds great promise in the realm of TSF, offering a prospective solution for large-scale, secure, and efficient time series prediction and analysis.

5.1.2 Lack of Interpretability

So far, the majority of efforts in the field of TSF have primarily focused on enhancing predictive performance through the design of intricate model architectures. However, research into the interpretability of these models has been relatively limited. As neural networks find application in critical tasks (Moraffah et al., 2020), the demand for comprehending why and how models make specific predictions has been growing. The N-BEATS model achieves high accuracy and interpretability in TSF by designing the interpretable architecture and output mechanisms (Oreshkin et al., 2019). This enables users to better comprehend the model's predictive outcomes while maintaining high forecasting precision.

迄今为止，时间序列预测（TSF）领域的研究主要集中于通过设计复杂模型架构来提升预测性能。然而，关于这些模型可解释性的研究仍相对有限。随着神经网络在关键任务中的应用（Moraffah等，2020），理解模型为何及如何做出特定预测的需求日益增长。N-BEATS模型通过设计可解释的架构与输出机制（Oreshkin等，2019），在保持高预测精度的同时实现了优异的可解释性，使用户能够更好地理解模型的预测结果

Post-hoc interpretable models are developed for the purpose of elucidating already trained networks, aiding in the identification of crucial features or instances without modifying the original model weights. These approaches mainly fall into two categories. One involves the application of simpler interpretable surrogate models between the inputs and outputs of the neural network, relying on these approximate models to provide explanations (Ribeiro et al., 2016; Lundberg and Lee, 2017). The other category encompasses gradient-based methods, such as those presented in (Simonyan et al., 2013; Koh and Liang, 2017; Siddiqui et al., 2019), which scrutinize the network gradients to determine which input features exert the most significant influence on the loss function.

事后可解释模型的开发旨在阐明已训练好的网络，帮助识别关键特征或实例，而无需修改原始模型权重。这些方法主要分为两大类：一类是在神经网络的输入与输出之间应用更简单的可解释替代模型，依靠这些近似模型提供解释（Ribeiro等，2016；Lundberg和Lee，2017）；另一类则包含基于梯度的方法，如（Simonyan等，2013；Koh和Liang，2017；Siddiqui等，2019）所提出的方法，通过分析网络梯度来确定哪些输入特征对损失函数具有最显著的影响

Furthermore, it is noteworthy that, in contrast to the black-box nature of traditional neural networks, a series of TSF models based on the Transformer architecture incorporate attention layers with inherent interpretability. These attention layers can be strategically integrated into other models, with the analysis of attention weights aiding in the comprehension of the relative importance of features at each time step (Choi et al., 2016; Li et al., 2019a; Bai et al., 2018b). By scrutinizing the distribution of attention vectors across time intervals, the model can gain better insights into persistent patterns or relationships within the time series (Lim et al., 2021), such as seasonal patterns.

此外值得注意的是，与传统神经网络的"黑箱"特性形成鲜明对比的是，基于Transformer架构的一系列时间序列预测模型采用了具有内在可解释性的注意力层。这些注意力层可被策略性地整合到其他模型中，通过分析注意力权重有助于理解每个时间步长中特征的相对重要性（Choi等，2016；Li等，2019a；Bai等，2018b）。通过细致考察注意力向量在不同时间区间的分布规律，模型能够更深入地把握时间序列中持续存在的模式或关联关系（Lim等，2021），例如季节性规律等

Recent advancements in the field have focused on learning from perturbations and interpretable sparse system identification methods to enhance the interpretability of time series data (Enguehard, 2023; Liu et al., 2024). Among these, sparse optimization methods, which obviate the need for time-consuming backpropagation training, exhibit efficient training capabilities on CPUs. These methods offer insights for further exploration into interpretable time series forecasting.

该领域的最新进展集中于通过扰动学习和可解释稀疏系统辨识方法提升时间序列数据的可解释性（Enguehard，2023；Liu等，2024）。其中，稀疏优化方法无需耗时的反向传播训练，在CPU上展现出高效的训练能力，这为深入探索可解释时间序列预测提供了新的研究路径

5.1.3 Lack of Temporal Continuity

Compared to traditional deep learning forecasting models, the proposal of the Neural Ordinary Differential Equation (NODE) (Chen et al., 2018) has directed our attention

相较于传统的深度学习预测模型，神经常微分方程（Neural Ordinary Differential Equation, NODE）（Chen等，2018）的提出将研究焦点引向神经网络参数化隐藏状态的导数层面。该模型在连续与离散时间序列问题上均展现出优于循环神经网络的性能。近期将常微分方程（ODE）或偏微分方程（PDE）应用于时间序列预测的研究探索了多个方向：包括学习变量或事件间的潜在关联（Li等，2022a；De Brouwer等，2019；Gao等，2022b）、处理不规则数据（Scholz等，2022）、实现可解释的连续性（Gao等，2022a；Jin等，2022）、优化模型参数（Chen等，2011）以及探究微分动力学（Linot等，2023；Gilani，2021）。Gao等（2022a）提出的ETN-ODE模型是首个具有可解释性的连续神经网络，可实现多变量在任意时间点的多步时间序列预测。此外，其EgPDE-Net模型（Gao等，2022b）亦首创性地将多元时间序列的连续时间表征构建为偏微分方程问题，通过特殊设计的架构利用ODE求解器将偏微分方程问题转化为常微分方程问题，从而支持任意时间步长的预测

towards the derivatives of neural network parameterized hidden states, which showcases superior performance over RNNs in both continuous and discrete time series problems. Recent studies applying Ordinary Differential Equations (ODE) or Partial Differential Equations (PDE) to TSF have explored various directions such as learning latent relationships between variables or events (Li et al., 2022a; De Brouwer et al., 2019; Gao et al., 2022b), handling irregular data (Scholz et al., 2022), achieving interpretable continuity (Gao et al., 2022a; Jin et al., 2022), optimizing model parameters (Chen et al., 2011), and exploring differential dynamics (Linot et al., 2023; Gilani, 2021). The ETN-ODE model proposed by Gao et al. (2022a) is the first interpretable continuous neural network for multi-step time series forecasting of multiple variables at arbitrary time instances. Additionally, their EgPDE-Net model (Gao et al., 2022b) is also the first to establish the continuous-time representation of multivariate time series as a partial differential equation problem. Its specially designed architecture utilizes ODE solvers to transform the partial differential equation problem into an ODE problem, facilitating predictions at arbitrary time steps.

时间连续性是可解释时间序列预测(TSF)过程中需要考虑的关键因素之一。神经微分方程(NDE)范式在可解释时间序列预测中的应用，通过将深度学习与微分方程建模相结合，能够自然而精确地捕捉时间序列的动态演化过程。该范式通过微分方程描述每个时间点数据的变化速率，从而更清晰地解释各组成要素的演化机制，并灵活捕捉瞬时变化。对于复杂时间序列数据的深度学习建模而言，神经微分方程技术提供了一种创新且高效的研究范式

Temporal continuation is one of the crucial factors to consider in the TSF process. The application of the Neural Differential Equation (NDE) paradigm in DTSF integrates DL with differential equation modeling to naturally and accurately capture the dynamic evolution of time series. It interprets the evolution of individual components more clearly and flexibly captures instantaneous changes by using a differential equation to describe the rate of change of the data at each time point. For deep learning modelling of complicated time series data, the NDE technique offers an innovative and effective paradigm.

5.1.4 Challenges of Parallel Computing

在大数据时代，对时间序列数据进行在线实时分析的需求日益迫切。当前时间序列模型的构建主要基于单机序列分析，通常需要依赖高性能GPU服务器来提升计算效率。然而一方面受限于计算资源和数据规模，难以实现实时在线预测；另一方面GPU服务器成本高昂。因此，基于深度学习与大数据分析技术的高效并行计算研究，将成为亟待攻克的关键难题

In the era of massive data, there is an urgent demand for online real-time analysis of time series data. Currently, time series models are constructed based on stand-alone sequence analysis, which often requires the use of high-performance GPU servers to improve computational efficiency. However, on one hand, it is constrained by computational resources and data scale, making real-time online forecasting unattainable. On the other hand, GPU servers are costly. Therefore, the research on efficient parallel computing based on deep learning and big data analytics technologies is poised to become a critical challenge.

5.1.5 Challenges of Large Models

大模型在时间序列预测领域展现出显著优势，其卓越性能体现在捕捉长期依赖关系、处理高维数据及抑制噪声等方面。2023年12月13日，亚马逊发布了利用大模型进行时间序列预测的研究成果（Xue and Salim, 2023），这标志着大模型在时序预测应用的开拓性探索。该研究通过大模型构建序列间复杂关联关系，同时充分发挥其强大的文本数据处理能力。大模型的引入强化了金融预测场景中多模态数据处理能力与模型可解释性。目前大模型已涉足多个专业领域，包括金融市场股价预测（Zhou et al, 2023; Jin et al, 2023; Chang et al, 2023）、医疗数据推理（Sun et al, 2023; Gruver et al, 2023）、人类移动轨迹预测（Cao et al, 2023a），以及作为通用模型应用于气象与能源需求预测（Yu et al, 2023; Xie et al, 2023; Zhang et al, 2023a; Liu et al, 2023b; Li et al, 2023a）

Large models demonstrate advantages in the field of time series forecasting, excelling in capturing long-term dependencies, handling high-dimensional data, and mitigating noise. A noteworthy exploration in this direction occurred on December 13, 2023 when Amazon released work utilizing large models for time series forecasting, marking a pioneering effort in applying large models to temporal prediction (Xue and Salim, 2023). This work leverages large models to construct intricate relationships between sequences while harnessing their robust text data processing capabilities. The integration of large models has enhanced the handling of multimodal data and interpretability in financial forecasting scenarios. Large models have already ventured into various domains, encompassing stock price predictions in financial markets (Zhou et al., 2023;

Jin et al., 2023; Chang et al., 2023), inference of medical data (Sun et al., 2023; Gruver et al., 2023), forecasting human mobility trajectories (Cao et al., 2023a), and serving as general-purpose models for weather and energy demand predictions (Yu et al., 2023; Xie et al., 2023; Zhang et al., 2023a; Liu et al., 2023b; Li et al., 2023a).

On another note, significant strides have been made in the training of foundational time series models (Xue et al., 2022; Garza and Mergenthaler-Canseco, 2023). The recent TimeGPT-1 model (Rasul et al., 2023) applies the techniques and architecture underlying large language models (LLM) to the forecasting domain, successfully establishing the first foundational time series model capable of zero-shot inference. This breakthrough opens avenues for creating foundational models specifically tailored for time series forecasting.

另一方面，基础时间序列模型的训练已取得显著进展（Xue等人，2022；Garza与Mergenthaler-Canseco，2023）。最新发布的TimeGPT-1模型（Rasul等人，2023）将大语言模型（LLM）的技术架构应用于预测领域，成功构建了首个具备零样本推理能力的基础时间序列模型。这一突破为开发专门针对时间序列预测的基础模型开辟了新路径

We believe that the performance and value of large models in the realm of time series forecasting will continue to unfold as technological advancements and innovations progress.

我们相信，随着技术进步与创新不断推进，大模型在时间序列预测领域的性能与价值将持续显现

5.2 Prospects

5.2.1 Potential Representation Learning

Representation Learning (RL) has recently emerged as one of the hot topics in time series forecasting. While models based on stacked layers can yield respectable results, they often come with high computational costs and may struggle to capture the inherent features of the data. RL, on the other hand, focuses on acquiring meaningful latent features that result in lower-dimensional and compact data representations, capturing the fundamental characteristics of the data. Presently, many self-supervised or unsupervised approaches aim to encode raw sequences to learn these latent representation features (Eldele et al., 2023; Darban et al., 2023). Some works employ multi-module architectures or model ensembles (Mehrkanoon, 2019; Lyu et al., 2018; Yang and Chen, 2019), while others use pre-training with denoising, smoothing properties, siamese structures or 2D-variation modeling (Zheng et al., 2023; Zerveas et al., 2021; Wu et al., 2022), which provide novel solutions to various domain-specific problems. Besides, contrastive learning is dedicated to enabling models to compare observations at different time points and learn rich data representations by contrasting positive and negative samples. Some works (Yue et al., 2022; Zhang et al., 2022a; Ozyurt et al., 2022; Luo et al., 2023) have utilized contrastive learning to assist models in learning meaningful features from unlabeled data, thus enhancing their generalization performance. This is especially valuable when labeled data is limited or unavailable.

表征学习（Representation Learning, RL）近年来已成为时间序列预测领域的研究热点之一。尽管基于堆叠层的模型能够取得可观的结果，但其通常伴随着高昂的计算成本，且难以捕捉数据的内在特征。相比之下，表征学习致力于获取有意义的潜在特征，从而生成低维紧凑的数据表征，以揭示数据的本质特性。当前，许多自监督或无监督方法旨在通过对原始序列进行编码来学习这些潜在表征特征（Eldele等，2023；Darban等，2023）。部分研究采用多模块架构或模型集成（Mehrkanoon，2019；Lyu等，2018；Yang和Chen，2019），另一些则利用去噪预训练、平滑特性、孪生结构或二维变分建模（Zheng等，2023；Zerveas等，2021；Wu等，2022），这些方法为特定领域问题提供了创新解决方案。此外，对比学习致力于使模型能够比较不同时间点的观测值，通过正负样本对比学习丰富的数据表征。部分研究成果（Yue等，2022；Zhang等，2022a；Ozyurt等，2022；Luo等，2023）已利用对比学习辅助模型从未标注数据中学习有意义的特征，从而提升其泛化性能——这一优势在标注数据有限或不可获取时显得尤为宝贵

Learning temporal representations and employing contrastive training can significantly enhance the model's representation and generalization capabilities in TSF. This greatly improves the model's performance in handling complex, noisy, or changing data distributions.

学习时序表征并采用对比训练能显著提升模型在时间序列预测中的表征能力和泛化性能。该方法可大幅增强模型处理复杂、含噪或动态变化数据分布时的表现

5.2.2 Counterfactual Forecast and Causal Inference

Counterfactual forecasting and causal inference represent promising avenues for future research in DTSF. Despite the existence of lots of deep learning methods for estimating causal effects in static settings (Yoon et al., 2018; Hartford et al., 2017; Alaa et al.,

反事实预测与因果推断为动态时间序列预测（DTSF）领域的未来研究提供了重要方向。尽管目前已存在诸多针对静态场景下因果效应估计的深度学习方法（Yoon等，2018；Hartford等，2017；Alaa等，2017），但时间序列数据面临的核心挑战在于时间依赖性混杂效应的存在。这一挑战源于时间序列的固有特性：影响目标变量的干预行为往往同时受到该目标变量历史观测值的制约。最新研究进展主要体现在统计技术的创新应用、新型损失函数的开发、现有方法的扩展改进以及适配的推理算法设计（Lim，2018；Bica等，2020；Li等，2020；Liu等，2023a；Gao等，2023）等方面

2017), the primary challenge in time series data lies in the presence of time-dependent confounding effects. This challenge arises due to the time-dependence, where actions that influence the target are also conditioned on observations of the target. Recent research advancements encompass the utilization of statistical techniques, novel loss functions, extensions of existing methods, and appropriate inference algorithms (Lim, 2018; Bica et al., 2020; Li et al., 2020; Liu et al., 2023a; Gao et al., 2023).

此外，尽管已有研究尝试为时间序列模型提供反事实解释（Dhaou等人，2021；Nemirovsky等人，2022），但这些方法在生成具有现实性或可行性的时间序列反事实解释方面仍存在不足。最新研究提出了一种可自解释的模型框架，该框架能够为时间序列预测任务生成可操作的反事实解释（Yan与Wang，2023）

Moreover, while some efforts provide counterfactual explanations for time series models (Dhaou et al., 2021; Nemirovsky et al., 2022), they fall short of generating realistic counterfactual explanations or feasible counterfactual explanations for time series models. Recent work has introduced a self-interpretable model capable of generating actionable counterfactual explanations for time series forecasting (Yan and Wang, 2023).

未来的研究方向可能围绕进一步完善这些方法展开，以解决时间序列数据固有的额外复杂性，并获得更准确的反事实解释。此外，应寻求创新方法来充分挖掘深度学习在反事实预测和因果推断中的潜力，最终提升各领域的决策流程

Future research directions may revolve around further refining these approaches to address the additional complexities inherent in time series data and get more accurate counterfactual interpretations. Additionally, innovative methods should be sought to harness the full potential of deep learning in counterfactual forecasting and causal inference, ultimately enhancing decision-making processes across various domains.

5.2.3 TS Diffusion

扩散模型在图像与视频流领域的蓬勃发展催生了诸多创新理论与模型，其应用疆域正逐步向时间序列预测（TSF）领域延伸。其中，TimeGrad通过循环神经网络引导的去噪过程实现自回归预测（Rasul等，2021），CSDI则采用自监督掩码技术的非自回归方法（Tashiro等，2021），而SSSD利用结构化状态空间模型以降低计算复杂度（Alcaraz与Strodthoff，2022）。尽管这些模型在时间序列预测领域属于早期探索，但仍存在推理速度迟缓、模型复杂度偏高以及边界一致性不足等固有缺陷

The burgeoning development of Diffusion models in the domain of image and video streams has sparked novel theories and models, gradually extending into the realm of TSF. Notably, TimeGrad employs RNN-guided denoising for autoregressive predictions (Rasul et al., 2021), while CSDI utilizes non-autoregressive methods with self-supervised masking (Tashiro et al., 2021). Similarly, SSSD utilizes structured state-space models to reduce computational complexity (Alcaraz and Strodthoff, 2022). Despite being early explorations in the TSF domain, these models still suffer from slow inference, high complexity, and boundary inconsistencies.

近期研究表明，无条件训练的TSDiff模型采用自引导机制，在不依赖辅助网络的情况下降低了逆向扩散过程在下游任务预测中的计算开销（Kollovieh等，2023）。TimeDiff模型通过未来混合技术与自回归初始化机制有效解决了边界不一致问题（Shen与Kwok，2023）。而多尺度扩散模型MR-Diff则利用多分辨率时间结构进行序列趋势提取与非自回归去噪（Shen等，2024）

In recent researches, the unconditionally trained TSDiff model employs self-guidance mechanisms to alleviate the computational overhead in reverse diffusion for downstream task forecasting without auxiliary networks (Kollovieh et al., 2023). TimeDiff addresses boundary inconsistencies with future mixups and autoregressive initialization mechanisms (Shen and Kwok, 2023). The multi-scale diffusion model MR-Diff leverages multi-resolution temporal structures for sequential trend extraction and non-autoregressive denoising (Shen et al., 2024).

首个基于DDPM的框架Diffusion-TS通过采用基于傅里叶变换的损失函数实现样本的精确重构，并将其扩展至预测任务领域（Yuan & Qiao，2024）。此外，TMDM模型将条件扩散生成过程与Transformer架构相结合，实现了多元时间序列的精确分布预测（Li等，2024b）

The first framework based on DDPM, Diffusion-TS, accurately reconstructs samples using Fourier-based loss functions, extending to forecasting tasks (Yuan and Qiao, 2024). Furthermore, the TMDM model combines conditional diffusion generation processes with Transformer to achieve precise distribution prediction for multivariate time series (Li et al., 2024b).

关于扩散模型的研究主要集中在去噪领域，同时在离散时间序列预测（DTSF）领域也涌现出诸多突破性成果。我们预期扩散模型将成为该领域的重要研究方向

The work on Diffusion primarily focuses on denoising, and numerous groundbreaking initiatives are emerging in the realm of DTSF. We anticipate Diffusion to become a prominent direction.

5.2.4 Determine the Weight of the Aggregate Model

At present, ensemble learning, as one of the mainstream paradigms, has proven to be effective and robust (Taylor et al., 2009; Makridakis et al., 2018b; Arbib, 2003). However, determining the weights of base models in an ensemble remains an unsolved challenge. Sub-optimal weighting can hinder the full potential of the final model. To address this challenge, Fu et al. (2022b) proposed a model combination framework based on reinforcement learning (RLMC). It uses deterministic policies to output dynamic model weights for non-stationary time series data and leverages deep learning to extract hidden features from raw time series data, allowing rapid adaptation to evolving data distributions. Notably, in RLMC, the use of DDPG, an off-policy actor-critic algorithm (Lillicrap et al., 2015), can produce continuous actions suitable for model combination problems and is trained with recorded data to achieve improved sample efficiency. Therefore, the combination of reinforcement learning with some continuous control algorithms (Fujimoto et al., 2018; Haarnoja et al., 2018) presents a unique utility in determining ensemble model weights and is a path worth exploring.

目前，集成学习作为主流范式之一，已被证明具有高效性与鲁棒性（Taylor等，2009；Makridakis等，2018b；Arbib，2003）。然而，如何确定基模型在集成中的权重仍是一个未解决的难题。次优的权重分配会制约最终模型的性能上限。针对这一挑战，Fu等（2022b）提出了基于强化学习的模型组合框架（RLMC）。该框架采用确定性策略输出非平稳时间序列数据的动态模型权重，并利用深度学习从原始时序数据中提取隐含特征，从而快速适应不断演变的数据分布。值得注意的是，RLMC中采用的DDPG算法——一种离轨执行的行动者-评论家算法（Lillicrap等，2015），能够生成适用于模型组合问题的连续动作，并通过记录数据进行训练以提高样本效率。因此，将强化学习与某些连续控制算法相结合（Fujimoto等，2018；Haarnoja等，2018），在确定集成模型权重方面展现出独特优势，是一条值得探索的研究路径

5.2.5 Interdisciplinary Exploration

Due to the multidimensional nature of the relationships between causes and effects in reality, there exist complex interconnections among time series. While deep learning models have demonstrated excellent performance in tackling intricate TSF problems, they often lack systematic interpretability and clear hierarchical structures. In the realm of network science, when dealing with extensive data, numerous variables, and intricate interconnections, it is possible to construct multi-layered networks by categorizing and stratifying the relationships among various elements. By examining the dynamic changes in multi-layered networks, it becomes feasible to forecast multidimensional data by analyzing high-dimensional correlations.

鉴于现实世界中因果关系具有多维特性，时间序列之间存在着复杂的相互关联。尽管深度学习模型在处理复杂时间序列预测问题时展现出卓越性能，但其往往缺乏系统化的可解释性与清晰的层次结构。在网络科学领域，当面对海量数据、多元变量及错综复杂的交互关系时，可通过要素间关系的分类与分层构建多层网络体系。通过观测多层网络的动态演变，借助高维关联性分析，即可实现对多维数据的预测

For diverse domains, an interdisciplinary approach, such as incorporating network science or other relevant theories, can be a beneficial choice in the future of DTSF research. This approach enables a more insightful analysis of problems and their multidimensional aspects.

针对不同领域，采用跨学科研究方法（如引入网络科学或其他相关理论）将成为未来动态交通信号配时（DTSF）研究中的有益选择。该方法有助于对问题及其多维特性进行更具洞察力的分析

6 Conclusion

In this paper, we present a systematic survey for deep learning-based time series forecasting. We commence with the fundamental definition of time series and forecasting tasks and summarize the statistical methods and their shortcomings. Next, moving on, we delve into neural network architectures for time series forecasting, summarizing five major model paradigms that have gained prominence in recent years: the Encoder-Decoder, Transformer, Generative Adversarial, Integration, and Cascade. Furthermore, we conduct an in-depth analysis of time series composition, elucidating the primary approaches to enhance feature extraction and learning from time series data. Additionally, we survey time series forecasting datasets across major domains, encompassing energy, healthcare, traffic, meteorology, and economics. Finally, we comprehensively outline the current challenges in the field and propose some potential research directions.

本文对基于深度学习的时间序列预测研究进行了系统性综述。首先从时间序列与预测任务的基本定义出发，梳理了传统统计方法及其局限性。随后重点探讨了时间序列预测的神经网络架构，归纳出近年来占据主导地位的五大模型范式：编码器-解码器架构、Transformer架构、生成对抗架构、集成架构以及级联架构。 进一步地，本文深入解析了时间序列的组成结构，阐明了增强时间序列特征提取与学习能力的主要技术路径。同时系统调研了能源、医疗、交通、气象、经济等主要领域的时间序列预测数据集。最后全面阐述了该领域当前面临的挑战，并提出了若干潜在的研究方向

Declarations

- **Funding**: This work was supported in part by the National Natural Science Foundation of China under Grant 62476247, 6207395 and 62072409, in part by the "Pioneer" and "Leading Goose" R&D Program of Zhejiang under Grant 2024C01214, and in part by the Zhejiang Provincial Natural Science Foundation under Grant LR21F020003.
- **Confict of interest**: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
- **Data availability**: Data sharing is not applicable to this article as no new data were created or analyzed in this study.

References

Ratnadip Adhikari and Ramesh K Agrawal. An introductory study on time series modeling and forecasting. *arXiv preprint arXiv:1302.6613*, 2013.

Nesreen K Ahmed, Amir F Atiya, Neamat El Gayar, and Hisham El-Shishiny. An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 29(5-6):594–621, 2010.

Ahmed M Alaa, Michael Weisz, and Mihaela Van Der Schaar. Deep counterfactual networks with propensity-dropout. *arXiv preprint arXiv:1706.05966*, 2017.

Juan Miguel Lopez Alcaraz and Nils Strodthoff. Diffusion-based time series imputation and forecasting with structured state space models. *Transactions on Machine Learning Research*, 2022. ISSN 2835-8856.

Jehad Ali, Rehanullah Khan, Nasir Ahmad, and Imran Maqsood. Random forests and decision trees. *International Journal of Computer Science Issues (IJCSI)*, 9(5):272, 2012.

Torben G Andersen, Tim Bollerslev, Peter Christoffersen, and Francis X Diebold. Volatility forecasting, 2005.

Abdul Fatir Ansari, Lorenzo Stella, Caner Turkmen, Xiyuan Zhang, Pedro Mercado, Huibin Shen, Oleksandr Shchur, Syama Sundar Rangapuram, Sebastian Pineda Arango, Shubham Kapoor, et al. Chronos: Learning the language of time series. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856.

Michael A Arbib. *The handbook of brain theory and neural networks.* MIT Press, 2003.

Mohammed Asiful, Rezaul Karim Hossain, Ruppa THulasiram, Neil DB Bruce, and Yang Wang. Hybrid deep learning model for stock price prediction. In *IEEE Symposium Series on Computational Intelligence, SSCI, Bangalore, India*, 2018.

Shaojie Bai, J Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*, 2018a.

Tian Bai, Shanshan Zhang, Brian L Egleston, and Slobodan Vucetic. Interpretable representation learning for healthcare via capturing disease progression through time. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 43–51, 2018b.

Kasun Bandara, Peibei Shi, Christoph Bergmeir, Hansika Hewamalage, Quoc Tran, and Brian Seaman. Sales demand forecast in e-commerce using a long short-term memory neural network methodology. In *Neural Information Processing: 26th International Conference, ICONIP 2019, Sydney, NSW, Australia, December 12–15, 2019, Proceedings, Part III 26*, pages 462–474. Springer, 2019.

Kasun Bandara, Christoph Bergmeir, and Hansika Hewamalage. Lstm-msnet: Leveraging forecasts on sets of related time series with multiple seasonal patterns. *IEEE Transactions on Neural Networks and Learning Systems*, 32(4):1586–1599, 2020.

Konstantinos Benidis, Syama Sundar Rangapuram, Valentin Flunkert, Yuyang Wang, Danielle Maddix, Caner Turkmen, Jan Gasthaus, Michael Bohlke-Schneider, David Salinas, Lorenzo Stella, et al. Deep learning for time series forecasting: Tutorial and literature survey. *ACM Computing Surveys*, 55(6):1–36, 2022.

David Berthelot, Rebecca Roelofs, Kihyuk Sohn, Nicholas Carlini, and Alexey Kurakin. Adamatch: A unified approach to semi-supervised learning and domain adaptation. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=Q5uh1Nvv5dm.

Hongjing Bi, Lilei Lu, and Yizhen Meng. Hierarchical attention network for multivariate time series long-term forecasting. *Applied Intelligence*, 53(5):5060–5071, 2023.

Ioana Bica, Ahmed M Alaa, James Jordon, and Mihaela van der Schaar. Estimating counterfactual treatment outcomes over time through adversarially balanced representations. *International Conference on Learning Representations*, 2020.

Joos-Hendrik Böse, Valentin Flunkert, Jan Gasthaus, Tim Januschowski, Dustin Lange, David Salinas, Sebastian Schelter, Matthias Seeger, and Yuyang Wang. Probabilistic demand forecasting at scale. *Proceedings of the VLDB Endowment*, 10(12): 1694–1705, 2017.

George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

Christopher Briggs, Zhong Fan, and Peter Andras. Federated learning for short-term residential load forecasting. *IEEE Open Access Journal of Power and Energy*, 9: 573–583, 2022.

Seok-Jun Bu and Sung-Bae Cho. Time series forecasting with multi-headed attention-based deep learning for residential energy consumption. *Energies*, 13(18):4722, 2020.

Ling Cai, Krzysztof Janowicz, Gengchen Mai, Bo Yan, and Rui Zhu. Traffic transformer: Capturing the continuity and periodicity of time series for traffic forecasting. *Transactions in GIS*, 24(3):736–755, 2020.

Zhaowei Cai and Nuno Vasconcelos. Cascade r-cnn: Delving into high quality object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6154–6162, 2018.

Laurent AF Callot, Anders B Kock, and Marcelo C Medeiros. Modeling and forecasting large realized covariance matrices and portfolio choice. *Journal of Applied Econometrics*, 32(1):140–158, 2017.

Defu Cao, Yujing Wang, Juanyong Duan, Ce Zhang, Xia Zhu, Congrui Huang, Yunhai Tong, Bixiong Xu, Jing Bai, Jie Tong, et al. Spectral temporal graph neural network for multivariate time-series forecasting. *Advances in Neural Information Processing*

Systems, 33:17766–17778, 2020.

Defu Cao, Furong Jia, Sercan O Arik, Tomas Pfister, Yixiang Zheng, Wen Ye, and Yan Liu. Tempo: Prompt-based generative pre-trained transformer for time series forecasting. *arXiv preprint arXiv:2310.04948*, 2023a.

Haizhou Cao, Zhenhao Huang, Tiechui Yao, Jue Wang, Hui He, and Yangang Wang. Inparformer: evolutionary decomposition transformers with interactive parallel attention for long-term time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023b.

Ching Chang, Wen-Chih Peng, and Tien-Fu Chen. Llm4ts: Two-stage fine-tuning for time-series forecasting with pre-trained llms. *arXiv preprint arXiv:2308.08469*, 2023.

Dunren Che, Mejdl Safran, and Zhiyong Peng. From big data to big data mining: challenges, issues, and opportunities. In *Database Systems for Advanced Applications: 18th International Conference, DASFAA 2013, International Workshops: BDMA, SNSM, SeCoP, Wuhan, China, April 22-25, 2013. Proceedings 18*, pages 1–15. Springer, 2013.

Chao Chen, Karl Petty, Alexander Skabardonis, Pravin Varaiya, and Zhanfeng Jia. Freeway performance measurement system: mining loop detector data. *Transportation Research Record*, 1748(1):96–102, 2001.

Guici Chen, Sijia Liu, and Feng Jiang. Daily weather forecasting based on deep learning model: A case study of shenzhen city, china. *Atmosphere*, 13(8):1208, 2022.

Mu-Yen Chen and Bo-Tsuen Chen. A hybrid fuzzy time series model based on granular computing for stock price forecasting. *Information Sciences*, 294:227–241, 2015.

Peng Chen, Yingying Zhang, Yunyao Cheng, Yang Shu, Yihang Wang, Qingsong Wen, Bin Yang, and Chenjuan Guo. Multi-scale transformers with adaptive pathways for time series forecasting. In *International Conference on Learning Representations*, 2024.

Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in Neural Information Processing Systems*, 31, 2018.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International Conference on Machine Learning*, pages 1597–1607. PMLR, 2020a.

Yitian Chen, Yanfei Kang, Yixiong Chen, and Zizhuo Wang. Probabilistic forecasting with temporal convolutional neural network. *Neurocomputing*, 399:491–501, 2020b.

Yuehui Chen, Bin Yang, Qingfang Meng, Yaou Zhao, and Ajith Abraham. Time-series forecasting using a system of ordinary differential equations. *Information Sciences*, 181(1):106–114, 2011.

Yushu Chen, Shengzhuo Liu, Jinzhe Yang, Hao Jing, Wenlai Zhao, and Guangwen Yang. A joint time-frequency domain transformer for multivariate time series forecasting. *arXiv preprint arXiv:2305.14649*, 2023.

Joseph Y Cheng, Hanlin Goh, Kaan Dogrusoz, Oncel Tuzel, and Erdrin Azemi. Subject-aware contrastive learning for biosignals. *arXiv preprint arXiv:2007.04871*, 2020.

Qi Cheng, Yixin Chen, Yuteng Xiao, Hongsheng Yin, and Weidong Liu. A dual-stage attention-based bi-lstm network for multivariate time series prediction. *The Journal of Supercomputing*, 78(14):16214–16235, 2022.

Kyunghyun Cho. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

Edward Choi, Mohammad Taha Bahadori, Jimeng Sun, Joshua Kulas, Andy Schuetz, and Walter Stewart. Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. *Advances in Neural Information Processing Systems*, 29, 2016.

Razvan-Gabriel Cirstea, Chenjuan Guo, Bin Yang, Tung Kieu, Xuanyi Dong, and Shirui Pan. Triformer: Triangular, variable-specific attentions for long sequence multivariate time series forecasting–full version. *arXiv preprint arXiv:2204.13767*, 2022.

Robert B Cleveland, William S Cleveland, Jean E McRae, and Irma Terpenning. Stl: A seasonal-trend decomposition. *J. Off. Stat*, 6(1):3–73, 1990.

John H Cochrane. Time series for macroeconomics and finance, 1997.

Zahra Zamanzadeh Darban, Geoffrey I Webb, Shirui Pan, and Mahsa Salehi. Carla: A self-supervised contrastive representation learning approach for time series anomaly detection. *arXiv preprint arXiv:2308.09296*, 2023.

Satyabrata Dash, Sujata Chakravarty, Sachi Nandan Mohanty, Chinmaya Ranjan Pattanaik, and Sarika Jain. A deep learning method to forecast covid-19 outbreak. *New Generation Computing*, 39(3-4):515–539, 2021.

Edward De Brouwer, Jaak Simm, Adam Arany, and Yves Moreau. Gru-ode-bayes: Continuous modeling of sporadically-observed time series. *Advances in Neural Information Processing Systems*, 32, 2019.

Alysha M De Livera, Rob J Hyndman, and Ralph D Snyder. Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American statistical association*, 106(496):1513–1527, 2011.

Chirag Deb, Fan Zhang, Junjing Yang, Siew Eang Lee, and Kwok Wei Shah. A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews*, 74:902–924, 2017.

Shumin Deng, Ningyu Zhang, Wen Zhang, Jiaoyan Chen, Jeff Z Pan, and Huajun Chen. Knowledge-driven stock trend prediction and explanation via temporal convolutional network. In *Companion Proceedings of the 2019 World Wide Web Conference*, pages 678–685, 2019.

Amin Dhaou, Antoine Bertoncello, Sébastien Gourvénec, Josselin Garnier, and Erwan Le Pennec. Causal and interpretable rules for time series analysis. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 2764–2772, 2021.

Fernando Díaz González. Federated learning for time series forecasting using lstm networks: Exploiting similarities through clustering, 2019.

Jiaxiang Dong, Haixu Wu, Haoran Zhang, Li Zhang, Jianmin Wang, and Mingsheng Long. Simmtm: A simple pre-training framework for masked time-series modeling. *arXiv preprint arXiv:2302.00861*, 2023.

Jiaxiang Dong, Haixu Wu, Yuxuan Wang, Yunzhong Qiu, Li Zhang, Jianmin Wang, and Mingsheng Long. Timesiam: A pre-training framework for siamese time-series modeling. *arXiv preprint arXiv:2402.02475*, 2024.

Alexey Dosovitskiy, Philipp Fischer, Jost Tobias Springenberg, Martin Riedmiller, and Thomas Brox. Discriminative unsupervised feature learning with exemplar convolutional neural networks (2015). *arXiv preprint arXiv:1406.6909*.

Alexandre Drouin, Étienne Marcotte, and Nicolas Chapados. Tactis: Transformer-attentional copulas for time series. In *International Conference on Machine Learning*, pages 5447–5493. PMLR, 2022.

Shengdong Du, Tianrui Li, Yan Yang, and Shi-Jinn Horng. Multivariate time series forecasting via attention-based encoder–decoder framework. *Neurocomputing*, 388: 269–279, 2020.

Wenying Duan, Xiaoxi He, Lu Zhou, Lothar Thiele, and Hong Rao. Combating distribution shift for accurate time series forecasting via hypernetworks. In *2022 IEEE 28th International Conference on Parallel and Distributed Systems (ICPADS)*, pages 900–907. IEEE, 2023.

Vijay Ekambaram, Arindam Jati, Nam Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. Tsmixer: Lightweight mlp-mixer model for multivariate time series forecasting. *arXiv preprint arXiv:2306.09364*, 2023.

Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, Chee Keong Kwoh, Xiaoli Li, and Cuntai Guan. Time-series representation learning via temporal and contextual contrasting. *arXiv preprint arXiv:2106.14112*, 2021.

Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, Chee-Keong Kwoh, Xiaoli Li, and Cuntai Guan. Self-supervised contrastive representation learning for semi-supervised time-series classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, and Xiaoli Li. Tslanet: Rethinking transformers for time series representation learning. *arXiv preprint arXiv:2404.08472*, 2024.

Joseph Enguehard. Learning perturbations to explain time series predictions. *arXiv preprint arXiv:2305.18840*, 2023.

Cristóbal Esteban, Stephanie L Hyland, and Gunnar Rätsch. Real-valued (medical) time series generation with recurrent conditional gans. *arXiv preprint arXiv:1706.02633*, 2017.

Christos Faloutsos, Jan Gasthaus, Tim Januschowski, and Yuyang Wang. Forecasting big time series: old and new. *Proceedings of the VLDB Endowment*, 11(12):2102–2105, 2018.

Christos Faloutsos, Valentin Flunkert, Jan Gasthaus, Tim Januschowski, and Yuyang Wang. Forecasting big time series: Theory and practice. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 3209–3210, 2019a.

Christos Faloutsos, Jan Gasthaus, Tim Januschowski, and Yuyang Wang. Classical and contemporary approaches to big time series forecasting. In *Proceedings of the 2019 International Conference on Management of Data*, pages 2042–2047, 2019b.

Jianqing Fan, Fang Han, and Han Liu. Challenges of big data analysis. *National Science Review*, 1(2):293–314, 2014.

Cong Feng, Erol Kevin Chartan, Bri-Mathias S Hodge, and Jie Zhang. Characterizing time series data diversity for wind forecasting. In *BDCAT*, pages 113–119, 2017.

Aya Ferchichi, Ali Ben Abbes, Vincent Barra, Manel Rhif, and Imed Riadh Farah. Multi-attention generative adversarial network for multi-step vegetation indices forecasting using multivariate time series. *Engineering Applications of Artificial Intelligence*, 128:107563, 2024.

Duarte Folgado, Marília Barandas, Ricardo Matias, Rodrigo Martins, Miguel Carvalho, and Hugo Gamboa. Time alignment measurement for time series. *Pattern Recognition*, 81:268–279, 2018.

En Fu, Yinong Zhang, Fan Yang, and Shuying Wang. Temporal self-attention-based conv-lstm network for multivariate time series prediction. *Neurocomputing*, 501: 162–173, 2022a.

Yuwei Fu, Di Wu, and Benoit Boulet. Reinforcement learning based dynamic model combination for time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022b.

Scott Fujimoto, Herke Hoof, and David Meger. Addressing function approximation error in actor-critic methods. In *International Conference on Machine Learning*, pages 1587–1596. PMLR, 2018.

Wayne A Fuller. *Introduction to statistical time series.* John Wiley & Sons, 2009.

Chris Funk, Pete Peterson, Martin Landsfeld, Diego Pedreros, James Verdin, Shraddhanand Shukla, Gregory Husak, James Rowland, Laura Harrison, Andrew Hoell, et al. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data*, 2(1):1–21, 2015.

Jiechao Gao, Wenpeng Wang, Zetian Liu, Md Fazlay Rabbi Masum Billah, and Bradford Campbell. Decentralized federated learning framework for the neighborhood: a case study on residential building load forecasting. In *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*, pages 453–459, 2021.

Penglei Gao, Xi Yang, Kaizhu Huang, Rui Zhang, and John Yannis Goulermas. Explainable tensorized neural ordinary differential equations for arbitrary-step time series prediction. *IEEE Transactions on Knowledge and Data Engineering*, 2022a.

Penglei Gao, Xi Yang, Kaizhu Huang, Rui Zhang, Ping Guo, and John Y Goulermas. Egpde-net: Building continuous neural networks for time series prediction with exogenous variables. *arXiv preprint arXiv:2208.01913*, 2022b.

Shanyun Gao, Raghavendra Addanki, Tong Yu, Ryan A. Rossi, and Murat Kocaoglu. Causal discovery in semi-stationary time series. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

Everette S Gardner Jr. Exponential smoothing: The state of the art. *Journal of Forecasting*, 4(1):1–28, 1985.

Azul Garza and Max Mergenthaler-Canseco. Timegpt-1. *arXiv preprint arXiv:2310.03589*, 2023.

Nigel Gebodh, Zeinab Esmaeilpour, Abhishek Datta, and Marom Bikson. Dataset of concurrent eeg, ecg, and behavior with multiple doses of transcranial electrical stimulation. *Scientific Data*, 8(1):274, 2021.

Faheem H Gilani. *Diffusion Maps and Its Applications to Time Series Forecasting and Filtering and Second Order Elliptic PDEs.* The Pennsylvania State University, 2021.

Rakshitha Godahewa, Christoph Bergmeir, Geoff Webb, Rob Hyndman, and Pablo Montero-Manso. Temperature rain dataset without missing values, 2021a.

Rakshitha Godahewa, Christoph Bergmeir, Geoff Webb, Pablo Montero-Manso, and Rob Hyndman. Dominick dataset, 2021b.

Zeying Gong, Yujin Tang, and Junwei Liang. Patchmixer: A patch-mixing architecture for long-term time series forecasting. *arXiv preprint arXiv:2310.00655*, 2023.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning.* MIT Press, 2016.

Nate Gruver, Marc Finzi, Shikai Qiu, and Andrew Gordon Wilson. Large language models are zero-shot time series forecasters. *arXiv preprint arXiv:2310.07820*, 2023.

Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. *arXiv preprint arXiv:2111.00396*, 2021.

Jing Guo, Penghui Lin, Limao Zhang, Yue Pan, and Zhonghua Xiao. Dynamic adaptive encoder-decoder deep learning networks for multivariate time series forecasting of building energy consumption. *Applied Energy*, 350:121803, 2023.

Na Guo, Cong Liu, Caihong Li, Qingtian Zeng, Chun Ouyang, Qingzhi Liu, and Xixi Lu. Explainable and effective process remaining time prediction using feature-informed cascade prediction model. *IEEE Transactions on Services Computing*, 2024.

Wenzhong Guo, Jianwen Wang, and Shiping Wang. Deep multimodal representation learning: A survey. *Ieee Access*, 7:63373–63394, 2019.

Tuomas Haarnoja, Aurick Zhou, Kristian Hartikainen, George Tucker, Sehoon Ha, Jie Tan, Vikash Kumar, Henry Zhu, Abhishek Gupta, Pieter Abbeel, et al. Soft actor-critic algorithms and applications. *arXiv preprint arXiv:1812.05905*, 2018.

James D Hamilton. *Time series analysis.* Princeton University Press, 2020.

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

Wenyong Han, Tao Zhu, Liming Chen, Huansheng Ning, Yang Luo, and Yaping Wan. Mcformer: Multivariate time series forecasting with mixed-channels transformer. *IEEE Internet of Things Journal*, 2024.

Jason Hartford, Greg Lewis, Kevin Leyton-Brown, and Matt Taddy. Deep iv: A flexible approach for counterfactual prediction. In *International Conference on Machine Learning*, pages 1414–1423. PMLR, 2017.

Andrew C Harvey. Forecasting, structural time series models and the kalman filter. 1990.

Xiaoyu He, Suixiang Shi, Xiulin Geng, and Lingyu Xu. Dynamic co-attention networks for multi-horizon forecasting in multivariate time series. *Future Generation Computer Systems*, 135:72–84, 2022.

Xiaoyu He, Suixiang Shi, Xiulin Geng, Jie Yu, and Lingyu Xu. Multi-step forecasting of multivariate time series using multi-attention collaborative network. *Expert Systems with Applications*, 211:118516, 2023.

Jeff Heaton. Ian goodfellow, yoshua bengio, and aaron courville: Deep learning: The mit press, 2016, 800 pp, isbn: 0262035618. *Genetic Programming and Evolvable Machines*, 19(1-2):305–307, 2018.

Keith W Hipel and A Ian McLeod. *Time series modelling of water resources and environmental systems*. Elsevier, 1994.

Ma Hiransha, E Ab Gopalakrishnan, Vijay Krishna Menon, and KP Soman. Nse stock market prediction using deep-learning models. *Procedia Computer Science*, 132:1351–1362, 2018.

Tin Kam Ho. Random decision forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition*, volume 1, pages 278–282. IEEE, 1995.

Jiaxi Hu, Disen Lan, Ziyu Zhou, Qingsong Wen, and Yuxuan Liang. Time-ssm: simplifying and unifying state space models for time series forecasting. *arXiv preprint arXiv:2405.16312*, 2024.

Rob J Hyndman and George Athanasopoulos. *Forecasting: Principles and Practice*. OTexts, 2018.

Romain Ilbert, Ambroise Odonnat, Vasilii Feofanov, Aladin Virmaux, Giuseppe Paolo, Themis Palpanas, and Ievgen Redko. Unlocking the potential of transformers in time series forecasting with sharpness-aware minimization and channel-wise attention. *arXiv preprint arXiv:2402.10198*, 2024.

Tim Januschowski, Jan Gasthaus, Yuyang Wang, David Salinas, Valentin Flunkert, Michael Bohlke-Schneider, and Laurent Callot. Criteria for classifying forecasting methods. *International Journal of Forecasting*, 36(1):167–177, 2020.

Paul Jeha, Michael Bohlke-Schneider, Pedro Mercado, Shubham Kapoor, Rajbir Singh Nirwan, Valentin Flunkert, Jan Gasthaus, and Tim Januschowski. Psa-gan: Progressive self attention gans for synthetic time series. 2022. The Tenth International Conference on Learning Representations, ICLR ; Conference date: 25-04-2022 Through 29-04-2022.

Ming Jin, Yu Zheng, Yuan-Fang Li, Siheng Chen, Bin Yang, and Shirui Pan. Multivariate time series forecasting with dynamic graph neural odes. *IEEE Transactions on Knowledge and Data Engineering*, 2022.

Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Y Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan, et al. Time-llm: Time series forecasting by reprogramming large language models. *arXiv preprint arXiv:2310.01728*, 2023.

Riya Kalra, Tinku Singh, Suryanshi Mishra, Naveen Kumar, Taehong Kim, Manish Kumar, et al. An efficient hybrid approach for forecasting real-time stock market indices. *Journal of King Saud University-Computer and Information Sciences*, 36 (8):102180, 2024.

Shruti Kaushik, Abhinav Choudhury, Pankaj Kumar Sheron, Nataraj Dasgupta, Sayee Natarajan, Larry A Pickett, and Varun Dutt. Ai in healthcare: time-series forecasting using statistical, neural, and ensemble architectures. *Frontiers in Big Data*, 3: 4, 2020.

Hendrik Klopries and Andreas Schwung. Itf-gan: Synthetic time series dataset generation and manipulation by interpretable features. *Knowledge-Based Systems*, 283: 111131, 2024.

Erdinc Koc and Muammer Türkoğlu. Forecasting of medical equipment demand and outbreak spreading based on deep long short-term memory network: the covid-19 pandemic in turkey. *Signal, Image and Video Processing*, pages 1–9, 2022.

Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deap: A database for emotion analysis; using physiological signals. *IEEE Transactions on Affective Computing*, 3(1):18–31, 2011.

Pang Wei Koh and Percy Liang. Understanding black-box predictions via influence functions. In *International Conference on Machine Learning*, pages 1885–1894. PMLR, 2017.

Stephan Kolassa. Why the "best" point forecast depends on the error or accuracy measure. *International Journal of Forecasting*, 36(1):208–211, 2020.

Marcel Kollovieh, Abdul Fatir Ansari, Michael Bohlke-Schneider, Jasper Zschiegner, Hao Wang, and Bernie Wang. Predict, refine, synthesize: Self-guiding diffusion models for probabilistic time series forecasting. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

Xiangjie Kong, Yuhan Wu, Hui Wang, and Feng Xia. Edge computing for internet of everything: A survey. *IEEE Internet of Things Journal*, 9(23):23472–23485, 2022.

Xiangjie Kong, Zhehui Shen, Kailai Wang, Guojiang Shen, and Yanjie Fu. Exploring bus stop mobility pattern: a multi-pattern deep learning prediction framework. *IEEE Transactions on Intelligent Transportation Systems*, 2024.

Peter Kontschieder, Madalina Fiterau, Antonio Criminisi, and Samuel Rota Bulo. Deep neural decision forests. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1467–1475, 2015.

Alireza Koochali, Peter Schichtel, Andreas Dengel, and Sheraz Ahmed. Probabilistic forecasting of sensory data with generative adversarial networks–forgan. *IEEE Access*, 7:63868–63880, 2019.

Nikolaos Kourentzes, Fotios Petropoulos, and Juan R Trapero. Improving forecasting by estimating time series structural components across multiple frequencies. *International Journal of Forecasting*, 30(2):291–302, 2014.

Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. Modeling long-and short-term temporal patterns with deep neural networks. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 95–104, 2018.

Pedro Lara-Benítez, Manuel Carranza-García, José M Luna-Romera, and José C Riquelme. Temporal convolutional networks applied to energy-related time series forecasting. *applied sciences*, 10(7):2322, 2020.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553): 436–444, 2015.

Junsoo Lee. Univariate time series modeling and forecasting (box-jenkins method). *Economic Times*, 413, 1994.

Sang Il Lee and Seong Joon Yoo. Threshold-based portfolio: the role of the threshold and its applications. *The Journal of Supercomputing*, 76(10):8040–8057, 2020.

Jun Li, Che Liu, Sibo Cheng, Rossella Arcucci, and Shenda Hong. Frozen language model helps ecg zero-shot learning. *arXiv preprint arXiv:2303.12311*, 2023a.

Longyuan Li, Junchi Yan, Yunhao Zhang, Jihai Zhang, Jie Bao, Yaohui Jin, and Xiaokang Yang. Learning generative rnn-ode for collaborative time-series and event sequence forecasting. *IEEE Transactions on Knowledge and Data Engineering*, 2022a.

Rui Li, Zach Shahn, Jun Li, Mingyu Lu, Prithwish Chakraborty, Daby Sow, Mohamed Ghalwash, and Li-wei H Lehman. G-net: a deep learning approach to g-computation for counterfactual outcome prediction under dynamic treatment regimes. *arXiv preprint arXiv:2003.10551*, 2020.

Shancang Li, Li Da Xu, and Shanshan Zhao. The internet of things: a survey. *Information Systems Frontiers*, 17:243–259, 2015.

Shiyang Li, Xiaoyong Jin, Yao Xuan, Xiyou Zhou, Wenhu Chen, Yu-Xiang Wang, and Xifeng Yan. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. *Advances in Neural Information Processing Systems*, 32, 2019a.

Tong Li, Zhaoyang Liu, Yanyan Shen, Xue Wang, Haokun Chen, and Sen Huang. Master: Market-guided stock transformer for stock price forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 162–170, 2024a.

Xuerong Li, Wei Shang, and Shouyang Wang. Text-based crude oil price forecasting: A deep learning approach. *International Journal of Forecasting*, 35(4):1548–1560, 2019b.

Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv preprint arXiv:1707.01926*, 2017.

Yan Li, Xinjiang Lu, Yaqing Wang, and Dejing Dou. Generative time series forecasting with diffusion, denoise, and disentanglement. *Advances in Neural Information Processing Systems*, 35:23009–23022, 2022b.

Yuan Li, Huanjie Wang, Jingwei Li, Chengbao Liu, and Jie Tan. Act: Adversarial convolutional transformer for time series forecasting. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2022c.

Yuxin Li, Wenchao Chen, Xinyue Hu, Bo Chen, Mingyuan Zhou, et al. Transformer-modulated diffusion models for probabilistic multivariate time series forecasting. In *The Twelfth International Conference on Learning Representations*, 2024b.

Zhe Li, Zhongwen Rao, Lujia Pan, Pengyun Wang, and Zenglin Xu. Ti-mae: Self-supervised masked time series autoencoders. *arXiv preprint arXiv:2301.08871*, 2023b.

Timothy P Lillicrap, Jonathan J Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*, 2015.

Bryan Lim. Forecasting treatment responses over time using recurrent marginal structural networks. *Advances in Neural Information Processing Systems*, 31, 2018.

Bryan Lim, Sercan Ö Arık, Nicolas Loeff, and Tomas Pfister. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4):1748–1764, 2021.

Shengsheng Lin, Weiwei Lin, Wentai Wu, Songbo Wang, and Yongxiang Wang. Petformer: Long-term time series forecasting via placeholder-enhanced transformer. *arXiv preprint arXiv:2308.04791*, 2023.

Wen-Hui Lin, Ping Wang, Kuo-Ming Chao, Hsiao-Chung Lin, Zong-Yu Yang, and Yu-Huang Lai. Wind power forecasting with deep learning networks: Time-series forecasting. *Applied Sciences*, 11(21):10335, 2021.

Alec J Linot, Joshua W Burby, Qi Tang, Prasanna Balaprakash, Michael D Graham, and Romit Maulik. Stabilized neural ordinary differential equations for long-time forecasting of dynamical systems. *Journal of Computational Physics*, 474:111838, 2023.

Mingzhou Liu, Xinwei Sun, Lingjing Hu, and Yizhou Wang. Causal discovery from subsampled time series with proxy variables. *arXiv preprint arXiv:2305.05276*, 2023a.

Minhao Liu, Ailing Zeng, Muxi Chen, Zhijian Xu, Qiuxia Lai, Lingna Ma, and Qiang Xu. Scinet: Time series modeling and forecasting with sample convolution and interaction. *Advances in Neural Information Processing Systems*, 35:5816–5828, 2022a.

Shizhan Liu, Hang Yu, Cong Liao, Jianguo Li, Weiyao Lin, Alex X Liu, and Schahram Dustdar. Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting. In *International Conference on Learning Representations*, 2021.

Xiaoyi Liu, Duxin Chen, Wenjia Wei, Xia Zhu, and Wenwu Yu. Interpretable sparse system identification: Beyond recent deep learning techniques on time-series prediction. In *The Twelfth International Conference on Learning Representations*, 2024.

Xin Liu, Daniel McDuff, Geza Kovacs, Isaac Galatzer-Levy, Jacob Sunshine, Jiening Zhan, Ming-Zher Poh, Shun Liao, Paolo Di Achille, and Shwetak Patel. Large language models are few-shot health learners. *arXiv preprint arXiv:2305.15525*, 2023b.

Yi Liu, JQ James, Jiawen Kang, Dusit Niyato, and Shuyu Zhang. Privacy-preserving traffic flow prediction: A federated learning approach. *IEEE Internet of Things Journal*, 7(8):7751–7763, 2020.

Yong Liu, Haixu Wu, Jianmin Wang, and Mingsheng Long. Non-stationary transformers: Exploring the stationarity in time series forecasting. *Advances in Neural Information Processing Systems*, 35:9881–9893, 2022b.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. *arXiv preprint arXiv:2310.06625*, 2023c.

Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 2017.

Dongsheng Luo, Wei Cheng, Yingheng Wang, Dongkuan Xu, Jingchao Ni, Wenchao Yu, Xuchao Zhang, Yanchi Liu, Yuncong Chen, Haifeng Chen, et al. Time series contrastive learning with information-aware augmentations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 4534–4542, 2023.

Rui Luo, Weinan Zhang, Xiaojun Xu, and Jun Wang. A neural stochastic volatility model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

Helmut Lütkepohl. Vector autoregressive moving average processes. *New Introduction to Multiple Time Series Analysis*, pages 419–446, 2005.

Yisheng Lv, Yanjie Duan, Wenwen Kang, Zhengxi Li, and Fei-Yue Wang. Traffic flow prediction with big data: A deep learning approach. *IEEE Transactions on*

Intelligent Transportation Systems, 16(2):865–873, 2014.

Xinrui Lyu, Matthias Hueser, Stephanie L Hyland, George Zerveas, and Gunnar Raetsch. Improving clinical predictions through unsupervised time series representation learning. *arXiv preprint arXiv:1812.00490*, 2018.

Spyros Makridakis. Time-series analysis and forecasting: An update and evaluation. *International Statistical Review/Revue Internationale de Statistique*, pages 255–278, 1978.

Spyros Makridakis, Evangelos Spiliotis, and Vassilios Assimakopoulos. The m4 competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*, 34(4):802–808, 2018a.

Spyros Makridakis, Evangelos Spiliotis, and Vassilios Assimakopoulos. Statistical and machine learning forecasting methods: Concerns and ways forward. *PloS one*, 13 (3):e0194889, 2018b.

Ricardo A Maronna, R Douglas Martin, Victor J Yohai, and Matías Salibián-Barrera. *Robust statistics: theory and methods (with R)*. John Wiley & Sons, 2019.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial Intelligence and Statistics*, pages 1273–1282. PMLR, 2017.

Siamak Mehrkanoon. Deep shared representation learning for weather elements forecasting. *Knowledge-Based Systems*, 179:120–128, 2019.

Suryanshi Mishra, Tinku Singh, Manish Kumar, and Satakshi. Multivariate time series short term forecasting using cumulative data of coronavirus. *Evolving Systems*, 15 (3):811–828, 2024.

Pablo Montero-Manso and Rob J Hyndman. Principles and algorithms for forecasting groups of time series: Locality and globality. *International Journal of Forecasting*, 37(4):1632–1653, 2021.

Douglas C Montgomery, Cheryl L Jennings, and Murat Kulahci. *Introduction to time series analysis and forecasting*. John Wiley & Sons, 2015.

Raha Moraffah, Mansooreh Karami, Ruocheng Guo, Adrienne Raglin, and Huan Liu. Causal interpretability for machine learning-problems, methods and evaluation. *ACM SIGKDD Explorations Newsletter*, 22(1):18–33, 2020.

Manfred Mudelsee. Trend analysis of climate time series: A review of methods. *Earth-science Reviews*, 190:310–322, 2019.

Daniel Nemirovsky, Nicolas Thiebaut, Ye Xu, and Abhishek Gupta. Countergan: Generating counterfactuals for real-time recourse and interpretability using residual gans. In *Uncertainty in Artificial Intelligence*, pages 1488–1497. PMLR, 2022.

Zelin Ni, Hang Yu, Shizhan Liu, Jianguo Li, and Weiyao Lin. Basisformer: Attention-based time series forecasting with learnable and interpretable basis. *arXiv preprint arXiv:2310.20496*, 2023.

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

Adamantios Ntakaris, Martin Magris, Juho Kanniainen, Moncef Gabbouj, and Alexandros Iosifidis. Benchmark dataset for mid-price forecasting of limit order book data with machine learning methods. *Journal of Forecasting*, 37(8):852–866, 2018.

Maureen O'Hara and Mao Ye. Is market fragmentation harming market quality? *Journal of Financial Economics*, 100(3):459–474, 2011.

Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*, 2016.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

Boris N Oreshkin, Dmitri Carpov, Nicolas Chapados, and Yoshua Bengio. N-beats: Neural basis expansion analysis for interpretable time series forecasting. *arXiv preprint arXiv:1905.10437*, 2019.

Ahmed Oussous, Fatima-Zahra Benjelloun, Ayoub Ait Lahcen, and Samir Belfkih. Big data technologies: A survey. *Journal of King Saud University-Computer and Information Sciences*, 30(4):431–448, 2018.

Yilmazcan Ozyurt, Stefan Feuerriegel, and Ce Zhang. Contrastive learning for unsupervised domain adaptation of time series. *arXiv preprint arXiv:2206.06243*, 2022.

Junwoo Park, Daehoon Gwak, Jaegul Choo, and Edward Choi. Self-supervised contrastive forecasting. In *The Twelfth International Conference on Learning Representations*, 2024.

Bo Peng, Yuanming Ding, and Wei Kang. Metaformer: A transformer that tends to mine metaphorical-level information. *Sensors*, 23(11):5093, 2023.

Hao Peng, Renyu Yang, Zheng Wang, Jianxin Li, Lifang He, S Yu Philip, Albert Y Zomaya, and Rajiv Ranjan. Lime: Low-cost and incremental learning for dynamic heterogeneous information networks. *IEEE Transactions on Computers*, 71(3):628–642, 2021.

Mathias Perslev, Michael Jensen, Sune Darkner, Poul Jørgen Jennum, and Christian Igel. U-time: A fully convolutional network for time series segmentation applied to sleep staging. *Advances in Neural Information Processing Systems*, 32, 2019.

Fotios Petropoulos, Daniele Apiletti, Vassilios Assimakopoulos, Mohamed Zied Babai, Devon K Barrow, Souhaib Ben Taieb, Christoph Bergmeir, Ricardo J Bessa, Jakub Bijak, John E Boylan, et al. Forecasting: Theory and practice. *International Journal of Forecasting*, 38(3):705–871, 2022.

Francesco Piccialli, Fabio Giampaolo, Edoardo Prezioso, David Camacho, and Giovanni Acampora. Artificial intelligence and healthcare: Forecasting of medical bookings through multi-source time-series fusion. *Information Fusion*, 74:1–16, 2021.

Yao Qin, Dongjin Song, Haifeng Chen, Wei Cheng, Guofei Jiang, and Garrison Cottrell. A dual-stage attention-based recurrent neural network for time series prediction. *arXiv preprint arXiv:1704.02971*, 2017.

Rial A Rajagukguk, Raden AA Ramadhan, and Hyun-Jin Lee. A review on deep learning models for forecasting time series data of solar irradiance and photovoltaic power. *Energies*, 13(24):6623, 2020.

Stephan Rasp, Peter D Dueben, Sebastian Scher, Jonathan A Weyn, Soukayna Mouatadid, and Nils Thuerey. Weatherbench: a benchmark data set for data-driven weather forecasting. *Journal of Advances in Modeling Earth Systems*, 12(11):

e2020MS002203, 2020.

Kashif Rasul, Calvin Seward, Ingmar Schuster, and Roland Vollgraf. Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In *International Conference on Machine Learning*, pages 8857–8868. PMLR, 2021.

Kashif Rasul, Arjun Ashok, Andrew Robert Williams, Arian Khorasani, George Adamopoulos, Rishika Bhagwatkar, Marin Biloš, Hena Ghonia, Nadhir Vincent Hassen, Anderson Schneider, et al. Lag-llama: Towards foundation models for time series forecasting. *arXiv preprint arXiv:2310.08278*, 2023.

Quentin Rebjock, Baris Kurt, Tim Januschowski, and Laurent Callot. Online false discovery rate control for anomaly detection in time series. *Advances in Neural Information Processing Systems*, 34:26487–26498, 2021.

Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1135–1144, 2016.

Lisbeth Rodríguez-Mazahua, Cristian-Aarón Rodríguez-Enríquez, José Luis Sánchez-Cervantes, Jair Cervantes, Jorge Luis García-Alcaraz, and Giner Alor-Hernández. A general perspective of big data: applications, tools, challenges and trends. *The Journal of Supercomputing*, 72:3073–3113, 2016.

Lior Rokach. Decision forest: Twenty years of research. *Information Fusion*, 27: 111–125, 2016.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

Frank Rosenblatt. *The perceptron, a perceiving and recognizing automaton Project Para*. Cornell Aeronautical Laboratory, 1957.

S Ruder. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*, 2017.

Seref Sagiroglu and Duygu Sinanc. Big data: A review. In *2013 International Conference on Collaboration Technologies and Systems (CTS)*, pages 42–47. IEEE, 2013.

David Salinas, Valentin Flunkert, Jan Gasthaus, and Tim Januschowski. Deepar: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3):1181–1191, 2020.

Pritam Sarkar and Ali Etemad. Self-supervised learning for ecg-based emotion recognition. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3217–3221. IEEE, 2020.

Marco Savi and Fabrizio Olivadese. Short-term energy consumption forecasting at the edge: A federated learning approach. *IEEE Access*, 9:95949–95969, 2021.

Harshit Saxena, Omar Aponte, and Katie T McConky. A hybrid machine learning model for forecasting a billing period's peak electric load days. *International Journal of Forecasting*, 35(4):1288–1303, 2019.

Randolf Scholz, Stefan Born, Nghia Duong-Trung, Mariano Nicolas Cruz-Bournazou, and Lars Schmidt-Thieme. Latent linear odes with neural kalman filtering for irregular time series forecasting. 2022.

Artemios-Anargyros Semenoglou, Evangelos Spiliotis, Spyros Makridakis, and Vassilios Assimakopoulos. Investigating the accuracy of cross-learning time series forecasting methods. *International Journal of Forecasting*, 37(3):1072–1084, 2021.

Ali Seyfi, Jean-Francois Rajotte, and Raymond Ng. Generating multivariate time series with common source coordinated gan (cosci-gan). *Advances in Neural Information Processing Systems*, 35:32777–32788, 2022.

Omer Berat Sezer, Mehmet Ugur Gudelek, and Ahmet Murat Ozbayoglu. Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Applied Soft Computing*, 90:106181, 2020.

Amin Shabani, Amir Abdi, Lili Meng, and Tristan Sylvain. Scaleformer: iterative multi-scale refining transformers for time series forecasting. *arXiv preprint arXiv:2206.04038*, 2022.

Zezhi Shao, Zhao Zhang, Fei Wang, and Yongjun Xu. Pre-training enhanced spatial-temporal graph neural network for multivariate time series forecasting. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1567–1577, 2022.

Lifeng Shen and James Kwok. Non-autoregressive conditional diffusion models for time series prediction. *arXiv preprint arXiv:2306.05043*, 2023.

Lifeng Shen, Weiyu Chen, and James Kwok. Multi-resolution diffusion models for time series forecasting. In *The Twelfth International Conference on Learning Representations*, 2024.

Wanxing Sheng, Keyan Liu, Dongli Jia, Shuo Chen, and Rongheng Lin. Short-term load forecasting algorithm based on lst-tcn in power distribution network. *Energies*, 15(15):5584, 2022.

Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. *Advances in Neural Information Processing Systems*, 28, 2015.

Robert H Shumway, David S Stoffer, and David S Stoffer. *Time series analysis and its applications*, volume 3. Springer, 2000.

Shoaib Ahmed Siddiqui, Dominique Mercier, Mohsin Munir, Andreas Dengel, and Sheraz Ahmed. Tsviz: Demystification of deep learning models for time-series analysis. *IEEE Access*, 7:67027–67040, 2019.

Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*, 2013.

Ankit Singh. Clda: Contrastive learning for semi-supervised domain adaptation. *Advances in Neural Information Processing Systems*, 34:5089–5101, 2021.

Shailendra Singh and Abdulsalam Yassine. Big data mining of energy time series for behavioral analytics and energy consumption forecasting. *Energies*, 11(2):452, 2018.

Tinku Singh, Nikhil Sharma, Satakshi, and Manish Kumar. Analysis and forecasting of air quality index based on satellite data. *Inhalation Toxicology*, 35(1-2):24–39, 2023.

Slawek Smyl. A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *International Journal of Forecasting*, 36(1):75–85, 2020.

Murat Cihan Sorkun, Christophe Paoli, and Özlem Durmaz Incel. Time series forecasting on solar irradiation using deep learning. In *2017 10th International Conference on Electrical and Electronics Engineering (ELECO)*, pages 151–155. IEEE, 2017.

Tejas Subramanya and Roberto Riggio. Centralized and federated learning for predictive vnf autoscaling in multi-domain 5g networks and beyond. *IEEE Transactions on Network and Service Management*, 18(1):63–78, 2021.

Chenxi Sun, Yaliang Li, Hongyan Li, and Shenda Hong. Test: Text prototype aligned embedding to activate llm's ability for time series. *arXiv preprint arXiv:2308.08241*, 2023.

Fan-Keng Sun and Duane S Boning. Fredo: frequency domain-based long-term time series forecasting. *arXiv preprint arXiv:2205.12301*, 2022.

Fan-Keng Sun, Chris Lang, and Duane Boning. Adjusting for autocorrelated errors in neural networks for time series. *Advances in Neural Information Processing Systems*, 34:29806–29819, 2021.

I Sutskever. Sequence to sequence learning with neural networks. *arXiv preprint arXiv:1409.3215*, 2014.

Afaf Taïk and Soumaya Cherkaoui. Electrical load forecasting using edge computing and federated learning. In *ICC 2020-2020 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE, 2020.

Shuntaro Takahashi, Yu Chen, and Kumiko Tanaka-Ishii. Modeling financial time-series with generative adversarial networks. *Physica A: Statistical Mechanics and its Applications*, 527:121261, 2019.

Peiwang Tang and Xianchao Zhang. Infomaxformer: Maximum entropy transformer for long time-series forecasting problem. *arXiv preprint arXiv:2301.01772*, 2023.

Yusuke Tashiro, Jiaming Song, Yang Song, and Stefano Ermon. Csdi: Conditional score-based diffusion models for probabilistic time series imputation. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 24804–24816. Curran Associates, Inc., 2021.

James W Taylor, Patrick E McSharry, and Roberto Buizza. Wind power density forecasting using ensemble predictions and time series models. *IEEE Transactions on Energy conversion*, 24(3):775–782, 2009.

David Alexander Tedjopurnomo, Zhifeng Bao, Baihua Zheng, Farhana Murtaza Choudhury, and Alex Kai Qin. A survey on modern deep neural network for traffic prediction: Trends, methods and challenges. *IEEE Transactions on Knowledge and Data Engineering*, 34(4):1544–1561, 2020.

Eric J Topol. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1):44–56, 2019.

José F Torres, Alicia Troncoso, Irena Koprinska, Zheng Wang, and Francisco Martínez-Álvarez. Deep learning for big data time series forecasting applied to solar power. In *International Joint Conference SOCO'18-CISIS'18-ICEUTE'18: San Sebastián, Spain, June 6-8, 2018 Proceedings 13*, pages 123–133. Springer, 2019.

Jean-François Toubeau, Jérémie Bottieau, François Vallée, and Zacharie De Grève. Deep learning-based multivariate probabilistic forecasting for short-term scheduling in power markets. *IEEE Transactions on Power Systems*, 34(2):1203–1215, 2018.

Artur Trindade. Electricityloaddiagrams20112014. UCI Machine Learning Repository, 2015.

Hao Wang and Zhenguo Zhang. Tatcn: time series prediction model based on time attention mechanism and tcn. In *2022 IEEE 2nd International Conference on Computer Communication and Artificial Intelligence (CCAI)*, pages 26–31. IEEE, 2022.

Lei Wang, Liang Zeng, and Jian Li. Aec-gan: adversarial error correction gans for auto-regressive long time-series generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 10140–10148, 2023.

Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma, James Y Zhang, and JUN ZHOU. Timemixer: Decomposable multiscale mixing for time series forecasting. In *International Conference on Learning Representations (ICLR)*, 2024a.

Xue Wang, Tian Zhou, Qingsong Wen, Jinyang Gao, Bolin Ding, and Rong Jin. Card: Channel aligned robust blend transformer for time series forecasting. In *The Twelfth International Conference on Learning Representations*, 2024b.

Yuxuan Wang, Haixu Wu, Jiaxiang Dong, Yong Liu, Yunzhong Qiu, Haoran Zhang, Jianmin Wang, and Mingsheng Long. Timexer: Empowering transformers for time series forecasting with exogenous variables. *arXiv preprint arXiv:2402.19072*, 2024c.

Zhiyuan Wang, Xovee Xu, Weifeng Zhang, Goce Trajcevski, Ting Zhong, and Fan Zhou. Learning latent seasonal-trend representations for time series forecasting. *Advances in Neural Information Processing Systems*, 35:38775–38787, 2022.

Ruofeng Wen, Kari Torkkola, Balakrishnan Narayanaswamy, and Dhruv Madeka. A multi-horizon quantile recurrent forecaster. *arXiv preprint arXiv:1711.11053*, 2017.

Gerald Woo, Chenghao Liu, Doyen Sahoo, Akshat Kumar, and Steven Hoi. Cost: Contrastive learning of disentangled seasonal-trend representations for time series forecasting. *arXiv preprint arXiv:2202.01575*, 2022a.

Gerald Woo, Chenghao Liu, Doyen Sahoo, Akshat Kumar, and Steven Hoi. Etsformer: Exponential smoothing transformers for time-series forecasting. *arXiv preprint arXiv:2202.01381*, 2022b.

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

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. *arXiv preprint arXiv:2210.02186*, 2022.

Neo Wu, Bradley Green, Xue Ben, and Shawn O'Banion. Deep transformer models for time series forecasting: The influenza prevalence case. *arXiv preprint arXiv:2001.08317*, 2020a.

Sifan Wu, Xi Xiao, Qianggang Ding, Peilin Zhao, Ying Wei, and Junzhou Huang. Adversarial sparse transformer for time series forecasting. *Advances in neural information processing systems*, 33:17105–17115, 2020b.

Xindong Wu, Xingquan Zhu, Gong-Qing Wu, and Wei Ding. Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*, 26(1):97–107, 2013.

Qianqian Xie, Weiguang Han, Yanzhao Lai, Min Peng, and Jimin Huang. The wall street neophyte: A zero-shot analysis of chatgpt over multimodal stock movement prediction challenges. *arXiv preprint arXiv:2304.05351*, 2023.

Qifa Xu, Xi Liu, Cuixia Jiang, and Keming Yu. Quantile autoregression neural network model with applications to evaluating value at risk. *Applied Soft Computing*, 49: 1–12, 2016.

Yingcheng Xu, Yunfeng Zhang, Peide Liu, Qiuyue Zhang, and Yuqi Zuo. Gan-enhanced nonlinear fusion model for stock price prediction. *International Journal of Computational Intelligence Systems*, 17(1):12, 2024.

Hao Xue and Flora D Salim. Promptcast: A new prompt-based learning paradigm for time series forecasting. *IEEE Transactions on Knowledge and Data Engineering*, pages 1–14, 2023.

Hao Xue, Bhanu Prakash Voutharoja, and Flora D Salim. Leveraging language foundation models for human mobility forecasting. In *Proceedings of the 30th International Conference on Advances in Geographic Information Systems*, pages 1–9, 2022.

Wang Xue, Tian Zhou, QingSong Wen, Jinyang Gao, Bolin Ding, and Rong Jin. Make transformer great again for time series forecasting: Channel aligned robust dual transformer. *arXiv preprint arXiv:2305.12095*, 2023.

Vijaya Krishna Yalavarthi, Kiran Madhusudhanan, Randolf Scholz, Nourhan Ahmed, Johannes Burchert, Shayan Jawed, Stefan Born, and Lars Schmidt-Thieme. Grafiti: Graphs for forecasting irregularly sampled time series. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 16255–16263, 2024.

Jingquan Yan and Hao Wang. Self-interpretable time series prediction with counterfactual explanations. *arXiv preprint arXiv:2306.06024*, 2023.

Hao-Fan Yang and Yi-Ping Phoebe Chen. Representation learning with extreme learning machines and empirical mode decomposition for wind speed forecasting methods. *Artificial Intelligence*, 277:103176, 2019.

Ye Yang and Jiangang Lu. A fusion transformer for multivariable time series forecasting: The mooney viscosity prediction case. *Entropy*, 24(4):528, 2022.

Kun Yi, Qi Zhang, Wei Fan, Shoujin Wang, Pengyang Wang, Hui He, Defu Lian, Ning An, Longbing Cao, and Zhendong Niu. Frequency-domain mlps are more effective learners in time series forecasting. *arXiv preprint arXiv:2311.06184*, 2023.

Ozge Cagcag Yolcu and Ufuk Yolcu. A novel intuitionistic fuzzy time series prediction model with cascaded structure for financial time series. *Expert Systems with Applications*, 215:119336, 2023.

Jinsung Yoon, James Jordon, and Mihaela Van Der Schaar. Ganite: Estimation of individualized treatment effects using generative adversarial nets. In *International Conference on Learning Representations*, 2018.

Jinsung Yoon, Daniel Jarrett, and Mihaela Van der Schaar. Time-series generative adversarial networks. *Advances in Neural Information Processing Systems*, 32, 2019.

Julong Young, Junhui Chen, Feihu Huang, and Jian Peng. Dateformer: Transformer extends look-back horizon to predict longer-term time series. 2022.

Bing Yu, Haoteng Yin, and Zhanxing Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv preprint arXiv:1709.04875*, 2017.

Chengqing Yu, Fei Wang, Zezhi Shao, Tangwen Qian, Zhao Zhang, Wei Wei, and Yongjun Xu. Ginar: An end-to-end multivariate time series forecasting model suitable for variable missing. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 3989–4000, 2024.

Xinli Yu, Zheng Chen, Yuan Ling, Shujing Dong, Zongyi Liu, and Yanbin Lu. Temporal data meets llm–explainable financial time series forecasting. *arXiv preprint arXiv:2306.11025*, 2023.

Xinyu Yuan and Yan Qiao. Diffusion-ts: Interpretable diffusion for general time series generation. In *The Twelfth International Conference on Learning Representations*, 2024.

Zhihan Yue, Yujing Wang, Juanyong Duan, Tianmeng Yang, Congrui Huang, Yunhai Tong, and Bixiong Xu. Ts2vec: Towards universal representation of time series. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 8980–8987, 2022.

Nur'atiah Zaini, Lee Woen Ean, Ali Najah Ahmed, and Marlinda Abdul Malek. A systematic literature review of deep learning neural network for time series air quality forecasting. *Environmental Science and Pollution Research*, pages 1–33, 2022.

Rizgar Zebari, Adnan Abdulazeez, Diyar Zeebaree, Dilovan Zebari, and Jwan Saeed. A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. *Journal of Applied Science and Technology Trends*, 1(1): 56–70, 2020.

George Zerveas, Srideepika Jayaraman, Dhaval Patel, Anuradha Bhamidipaty, and Carsten Eickhoff. A transformer-based framework for multivariate time series representation learning. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 2114–2124, 2021.

Boyu Zhang, Hongyang Yang, and Xiao-Yang Liu. Instruct-fingpt: Financial sentiment analysis by instruction tuning of general-purpose large language models. *arXiv preprint arXiv:2306.12659*, 2023a.

Chenhan Zhang, Shuyu Zhang, JQ James, and Shui Yu. Fastgnn: A topological information protected federated learning approach for traffic speed forecasting. *IEEE Transactions on Industrial Informatics*, 17(12):8464–8474, 2021.

G Peter Zhang. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159–175, 2003.

Guoqiang Zhang, B Eddy Patuwo, and Michael Y Hu. Forecasting with artificial neural networks:: The state of the art. *International Journal of Forecasting*, 14(1): 35–62, 1998.

Wenrui Zhang, Ling Yang, Shijia Geng, and Shenda Hong. Self-supervised time series representation learning via cross reconstruction transformer. *arXiv preprint arXiv:2205.09928*, 2022a.

Xiaoning Zhang, Fang Fang, and Jiaqi Wang. Probabilistic solar irradiation forecasting based on variational bayesian inference with secure federated learning. *IEEE Transactions on Industrial Informatics*, 17(11):7849–7859, 2020.

Xiyuan Zhang, Xiaoyong Jin, Karthick Gopalswamy, Gaurav Gupta, Youngsuk Park, Xingjian Shi, Hao Wang, Danielle C Maddix, and Yuyang Wang. First detrend then attend: Rethinking attention for time-series forecasting. *arXiv preprint arXiv:2212.08151*, 2022b.

Yifan Zhang, Rui Wu, Sergiu M Dascalu, and Frederick C Harris Jr. Multi-scale transformer pyramid networks for multivariate time series forecasting. *arXiv preprint arXiv:2308.11946*, 2023b.

Yunhao Zhang and Junchi Yan. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In *The eleventh international conference on learning representations*, 2023.

Zhenwei Zhang, Xin Wang, and Yuantao Gu. Sageformer: Series-aware graph-enhanced transformers for multivariate time series forecasting. *arXiv preprint arXiv:2307.01616*, 2023c.

Zhenwei Zhang, Linghang Meng, and Yuantao Gu. Sageformer: Series-aware framework for long-term multivariate time series forecasting. *IEEE Internet of Things Journal*, 2024. doi: 10.1109/JIOT.2024.3363451.

Wentian Zhao, Yanyun Gao, Tingxiang Ji, Xili Wan, Feng Ye, and Guangwei Bai. Deep temporal convolutional networks for short-term traffic flow forecasting. *Ieee Access*, 7:114496–114507, 2019.

Yongning Zhao, Lin Ye, Zhi Li, Xuri Song, Yansheng Lang, and Jian Su. A novel bidirectional mechanism based on time series model for wind power forecasting. *Applied Energy*, 177:793–803, 2016.

Xiaochen Zheng, Xingyu Chen, Manuel Schürch, Amina Mollaysa, Ahmed Allam, and Michael Krauthammer. Simts: Rethinking contrastive representation learning for time series forecasting. *arXiv preprint arXiv:2303.18205*, 2023.

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 11106–11115, 2021.

Tian Zhou, Ziqing Ma, Qingsong Wen, Liang Sun, Tao Yao, Wotao Yin, Rong Jin, et al. Film: Frequency improved legendre memory model for long-term time series forecasting. *Advances in Neural Information Processing Systems*, 35:12677–12690, 2022a.

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

Tian Zhou, Jianqing Zhu, Xue Wang, Ziqing Ma, Qingsong Wen, Liang Sun, and Rong Jin. Treedrnet: a robust deep model for long term time series forecasting. *arXiv preprint arXiv:2206.12106*, 2022c.

Tian Zhou, Peisong Niu, Xue Wang, Liang Sun, and Rong Jin. One fits all: Power general time series analysis by pretrained lm. *arXiv preprint arXiv:2302.11939*, 2023.

Hongjun Zhu, Shun Yuan, Xin Liu, Kuo Chen, Chaolong Jia, and Ying Qian. Casciff: A cross-domain information fusion framework tailored for cascade prediction in social networks. *Knowledge-Based Systems*, page 112391, 2024.

Yuzhen Zhu, Shaojie Luo, Di Huang, Weiyan Zheng, Fang Su, and Beiping Hou. Drcnn: decomposing residual convolutional neural networks for time series forecasting. *Scientific Reports*, 13(1):15901, 2023.

Dongcheng Zou, Senzhang Wang, Xuefeng Li, Hao Peng, Yuandong Wang, Chunyang Liu, Kehua Sheng, and Bo Zhang. Multispans: A multi-range spatial-temporal transformer network for traffic forecast via structural entropy optimization. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, pages 1032–1041, 2024.

A Datasets in Different Domain

Time series, which exists in every aspect of our lives, carries the historical data of various fields in the time dimension. Many datasets have been accumulated during the development of the TSF task. These datasets are often cited in top conferences and journals within the computer domain, furnishing researchers with high-quality research data characterized by rich samples and features, thus holding significant reference value. However, the diversity of these datasets introduces a significant challenge—data heterogeneity. The datasets described below cover five key TSF application areas: energy, transportation, economics, meteorology, and healthcare (Gebodh et al., 2021), as shown in Table 3. These fields feature data with varying structures, formats, time granularities, and scales, such as sensor data, text, and images, complicating model construction. To address these issues, several techniques have been proposed.

Multimodal learning, through shared representation learning, integrates diverse data types, improving model handling of heterogeneous data (Guo et al., 2019). Time alignment techniques, such as the TAM model, synchronize data from different time granularities by introducing a novel time-distance measure (Folgado et al., 2018). Deep generative models, like GinAR, address missing values and noise by generating new samples and rebuilding spatiotemporal dependencies (Yu et al., 2024). Self-supervised learning methods, such as SimCLR, allow models to learn from unlabeled data, improving adaptability to heterogeneous sources (Chen et al., 2020a). Finally, collaborative attention mechanisms capture complex correlations between multimodal data and adjust modality weights dynamically, enhancing model learning capacity (Dosovitskiy et al.). These models and techniques effectively integrate heterogeneous data, improving the stability and accuracy of time series forecasting in multi-source environments.

A.1 Energy

TSF is currently being extensively applied in a prominent domain, namely, energy management. Accurate forecasting within this domain plays a crucial role in facilitating status assessment and trend analysis, which in turn enables the implementation of intelligent strategies in engineering planning. Fortunately, modern energy systems autonomously gather extensive datasets encompassing diverse energy sources such as electricity (Singh and Yassine, 2018), wind energy (Feng et al., 2017), and solar energy (Rajagukguk et al., 2020). These data resources are leveraged for the identification of patterns and trends in energy demand and supply, providing valuable insights for the development of advanced forecasting models.

A.1.1 Electricity Transformer Temperature (ETT)

The ETT-small dataset encompasses data originating from two distinct power transformer installations, each situated at a separate site (Zhou et al., 2021). This dataset comprises a variety of parameters, such as load profiles and oil temperature readings. It serves the purpose of predicting the oil temperature of power transformers and investigating their resilience under extreme load conditions. The temporal scope of this dataset spans from July 2016 to July 2018, with data recorded at 15-minute

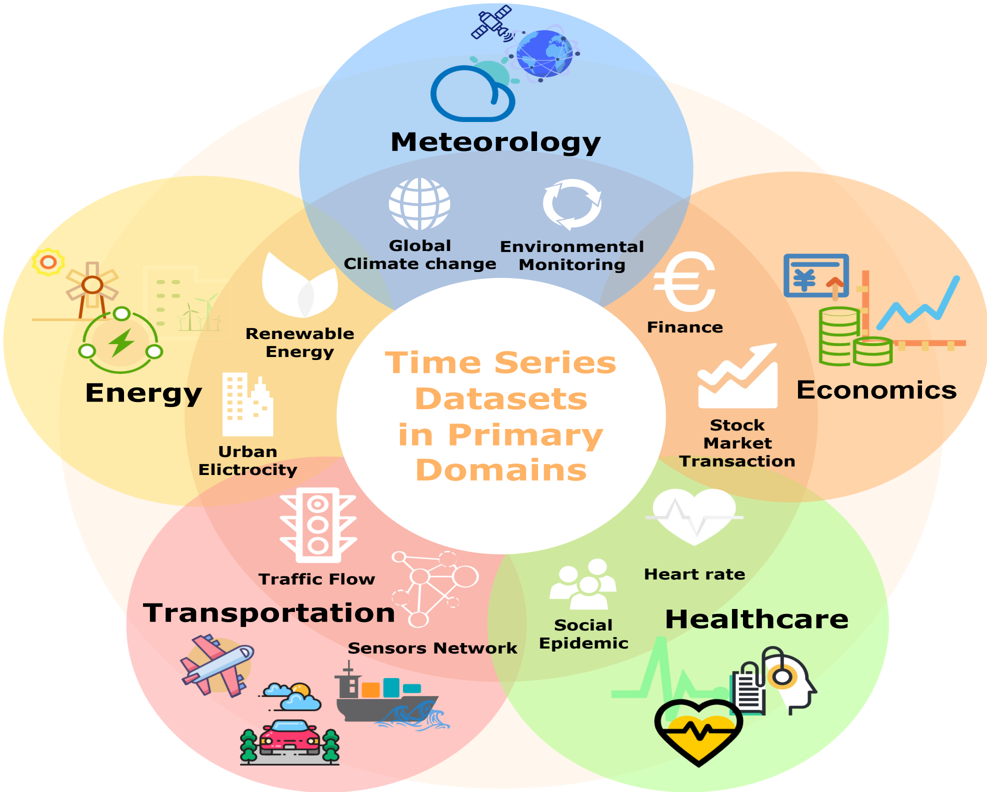


Fig. 16 Time series datasets in primary domains

intervals. These datasets originate from two geographically disparate regions within the same province in China, designated as ETT-small-m1 and ETT-small-m2, respectively. Each of these datasets consists of an extensive 70,080 data points, calculated based on a duration of 2 years, 365 days per year, 24 hours per day, and data sampling at 15-minute intervals. Furthermore, the dataset offers an alternate version with hourly granularity, denoted as ETT-small-h1 and ETT-small-h2. Each data point within the ETT dataset is characterized by an 8-dimensional feature vector, which includes the timestamp of the data point, the target variable 'oil temperature', and six distinct types of external load values.

A.1.2 Electricity

The initial dataset utilized in this investigation is the Electricity Load Diagrams 2011-2014 Dataset (Trindade, 2015), which records 370 customers' electricity usage information between 2011 and 2014. Data is recorded in the original dataset every 15 minutes. It was necessary to preprocess the dataset by deleting the 2011 data and aggregating it into hourly consumption in order to address the problem of some dimensions having a value of 0. As a result, the final dataset includes information on 321 customers' electrical use from 2012 to 2014.

A.1.3 Wind (European Wind Generation)

For 28 European countries between 1986 and 2015, this dataset [1] offers hourly estimates of energy potential expressed as a percentage of the maximum output from power plants. It is distinguished from other datasets by having sparser data and a notable frequency of zeros at regular intervals.

A.1.4 Solar-Energy

The solar power production of 137 photovoltaic plants in Alabama State in 2006, recorded at 10-minute intervals, constitutes the dataset for our evaluation of short-sequence forecasting capabilities [2].

A.2 Healthcare

TSF plays a pivotal role in the healthcare domain, serving as a critical tool for predicting disease onset and progression, evaluating the efficacy of pharmaceutical interventions, and monitoring fluctuations in patients' vital signs. These forecasts empower healthcare practitioners in enhancing disease diagnosis, devising treatment strategies, overseeing patient well-being, and implementing preventive measures for disease surveillance and containment.

A.2.1 ILI (Influenza-Like Illness)

Weekly reports from the US Centers for Disease Control and Prevention from 2002 to 2021 are included in the set of data. It contains data on the overall number of patients as well as the percentage of patients having influenza-like symptoms.

A.2.2 EEG (Electroencephalogram)

The collection includes EEG [3] recordings of participants obtained both prior to and during the performance of mental math exercises. Every recording is made up of 60-second EEG segments free of artifacts. For every subject in the dataset, there are 36 CSV files total, and each file has 19 data channels.

A.2.3 MIT-BIH (Arrhythmia Database)

There are 48 half-hour segments of two-channel ambulatory ECG recordings available in the MIT-BIH Arrhythmia Database [4]. These recordings were from 47 individuals that the BIH Arrhythmia Laboratory examined from 1975 to 1979. Every recording was digitalized with a resolution of 11 bits and a range of 10 mV, at a rate of 360

---

[1] https://www.kaggle.com/datasets/sohier/30-years-of-european-wind-generation
[2] https://www.nrel.gov/grid/solar-power-data.html
[3] https://github.com/meagmohit/EEG-Datasets
[4] http://ecg.mit.edu/

samples per second per channel. Electrocardiogram data from this dataset can be used for anticipating arrhythmias, among other uses.

A.3 Transportation

Accurate and timely TSF of traffic is vital for urban traffic control and management. It aids in predicting traffic congestion, traffic flow, accident rates, and the utilization of public transportation. These predictions can be used by transportation authorities and companies to plan and manage transportation systems more effectively, thereby improving traffic efficiency and safety.

A.3.1 Traffic

This dataset [5] includes hourly data from 2015–2016 that was collected during a 48-month period from the California Department of Transportation. The statistic shows the hourly road occupancy rate, which ranges from 0 to 1. The San Francisco Bay Area's roadways are home to 862 different sensors from which the measurements are obtained.

A.3.2 PeMSD4/7/8

These datasets are highly regarded as industry standards for traffic forecasting (Chen et al., 2001).

PeMSD4 is one of them and it includes traffic speed data from the San Francisco Bay Area. It incorporates data from 29 roads' worth of 307 sensors. The January–February 2018 time frame is covered by the dataset.

PeMSD7 includes traffic information from California's District 7. It covers the workday period from May to June 2012 and includes traffic speeds recorded by 228 sensors. Five minutes are allotted for the collection of data.

PeMSD8 contains San Bernardino traffic statistics taken during July and August of 2016. It includes data from 170 detectors positioned along 8 distinct routes. Five minutes are allotted for the collection of data.

A.4 Meteorology

TSF has become an indispensable task in the field of meteorology with wide-ranging applications in weather forecasting, such as meteorological disaster warnings, agricultural production, and more.

A.4.1 Weather1

The dataset Weather1 encompasses climate data from almost 1600 locations in the United States [6], spanning a 4-year period from 2010 to 2013. Hourly data points were collected, featuring the target value "wet bulb" and 11 climate-related features.

[5] https://pems.dot.ca.gov/
[6] https://www.ncei.noaa.gov/data/local-climatological-data/

A.4.2 Weather2

Weather2 comprises a meteorological time series featuring 21 weather indicators [7], collected every 10 minutes in 2020 by the Max Planck Institute for Biogeochemistry's weather station.

A.4.3 Temperature Rain

Consisting of 32,072 daily time series, this dataset (Godahewa et al., 2021a) presents temperature observations and rain forecasts collected by the Australian Bureau of Meteorology. The data spans 422 weather stations across Australia, covering the period from 02/05/2015 to 26/04/2017.

A.5 Economics

In the field of finance, one of the most extensively studied areas in TSF is the prediction of financial time series, particularly asset prices. Typically, there are several subtopics in this field, including stock price prediction, index prediction, foreign exchange price prediction, commodity (such as oil, gold, etc.) price prediction, bond price prediction, volatility prediction, and cryptocurrency price prediction. The following section will introduce commonly used datasets in this domain.

A.5.1 Exchange-Rate

This dataset (Lai et al., 2018) compiles daily exchange rates mainly in trading days for eight countries (Australia, Canada, China, Japan, New Zealand, Singapore, Switzerland, and the United Kingdom) spanning the years 1990 to 2016.

A.5.2 LOB-ITCH

Due to the lack of adequate records, few other fields have Millisecond data on the span of days as in finance. In the financial field, with the advent of automated trading, limit order books were born, which are very conducive to high-frequency traders' operations and leave a large amount of detailed data. The LOB-ITC dataset comprises around four million events, each with a 144-dimensional representation, pertaining over five stocks for ten consecutive trading days (Ntakaris et al., 2018), from June 1, 2010 to June 14, 2010. And what makes this data different from other data of the same kind is the centralized trading market in the Nordic region.Some researchers found that "the differences between different trading platforms' matching rules and transaction costs complicate comparisons between different limit order books for the same asset (O'Hara and Ye, 2011)" . Therefore, Stock Exchange, which has decentralized exchanges like the United States, has more influencing factors and is more difficult to model. In contrast, Helsinki Exchange is a pure limit order market, which can provide purer data.

[7] https://www.bgc-jena.mpg.de/wetter/

A.5.3 Dominick

This dataset (Godahewa et al., 2021b) incorporates data from randomized experiments conducted by the University of Chicago Booth School of Business and the now-defunct Dominick's Finer Foods. The experiments spanned from 1989 to 1994, covering over 25 different categories across all 100 stores in the chain. As a result of this research collaboration, approximately nine years of store-level data on the sales of more than 3,500 UPCs are available through this resource.

A.6 Further Data Sources

In addition to the commonly used datasets mentioned above, we extensively surveyed data sources from various domains and compiled a subset of additional datasets. These datasets are derived from influential works and serve as the foundation for researching niche topics and detailed investigations in respective fields. We will provide appropriate descriptions of the datasets listed in Table 4.

Several comprehensive datasets from large-scale competitions are also noteworthy, such as M3/M4/M5. These datasets were put forward by the Makridakis Competitions, which are a series of open competitions to evaluate and compare the accuracy of different TSF methods.

A.6.1 M3

This dataset [8] comprises yearly, quarterly, monthly, daily, and other time series. To ensure the development of accurate forecasting models, minimum observation thresholds were established: 14 for yearly series, 16 for quarterly series, 48 for monthly series, and 60 for other series. Time series within the domains of micro, industry, macro, finance, demographic, and others were included.

A.6.2 M4

The M4 dataset (Makridakis et al., 2018a) encompasses 100,000 real-life series in diverse domains, including micro, industry, macro, finance, demographic, and others.

A.6.3 M5

Covering stores in three US States (California, Texas, and Wisconsin), this dataset [9] includes item-level, department, product categories, and store details. It incorporates explanatory variables such as price, promotions, day of the week, and special events. Alongside time series data, it incorporates additional explanatory variables (e.g., Super Bowl, Valentine's Day, and Orthodox Easter) influencing sales, enhancing forecasting accuracy.

[8] https://forecasters.org/resources/time-series-data/
[9] https://mofc.unic.ac.cy/m5-competition/

Table 4 Summary of the datasets used in the experiments

Domain	Variants	Dataset	Data Time Range	Data Granularity	Reference
Energy	21	the Scada wind farm in Turkey	2018/1/1-2018/12/29	10m	(Lin et al., 2021)
	-	Global horizontal solar radiation data	1998/1/1-2007/12/1	1h	(Sorkun et al., 2017)
	-	Rooftop PV plant	2015/1/1-2016/12/31	30m	(Torres et al., 2019)
	9	UCI household electric power consumption	2006/12-2010/11	1m	(Bu and Cho, 2020)
	-	Spanish electricity demand	2014/01/02-2019/11/01	10m	(Lara-Benítez et al., 2020)
	-	Electric Vehicles Power Consumption	2015/3/2-2016/5/31	1h	(Lara-Benítez et al., 2020)
Healthcare	-	CDC ILI data	2010-2018	1d	(Wu et al., 2020a)
	45	DEAP	-	1 interval	(Koelstra et al., 2011)
	9	Turkish COVID-19 data	2020/3/27-2020/6/11	1d	(Koc and Türkoğlu, 2022)
	9	COVID-19 dataset of Orissa state	2020/1/30-2020/6/11	1d	(Dash et al., 2021)
Transportation	207	METR-LA	2012/3/1-2012/6/30	5m	(Cai et al., 2020)
	325	PeMS-BAY	2017/1/1-2017/5/31	5m	(Cai et al., 2020)
	-	BJER4	2014/7/1-2014/8/31	5m	(Yu et al., 2017)
Meteorology	6	Daily data of Shenzhen	from 2015	-	(Chen et al., 2022)
	-	CHIRPS	1981-2015	-	(Funk et al., 2015)
	-	WeatherBench	-	-	(Rasp et al., 2020)
Economics	5	S&P500	1997/1/1-2016/12/1	1d	(Lee and Yoo, 2020)
	13	NSE stocks data	1996/1/1-2015/6/30	1d	(Hiransha et al., 2018)
	6	NYSE stock data	2011/1/3-2016/12/30	1m	(Hiransha et al., 2018)

A.6.4 M6

The dataset [10] comprises two categories of assets: one selected from the Standard & Poor's 500 Index, consisting of 50 stocks, and the other comprising 50 Exchange-Traded Funds (ETFs) from various international exchanges. The focus of the M6 competition lies in forecasting the returns and risks associated with these stocks, along with investment decisions made based on the aforementioned predictions.

[10] https://mofc.unic.ac.cy/