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]

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.

✉ Feng Xia
f.xia@ieee.org

Xiangjie Kong
xjkong@ieee.org

Zhenghao Chen
202006010504@zjut.edu.cn

Weiyao Liu
211123120037@zjut.edu.cn

Kaili Ning
211124120018@zjut.edu.cn

Lechao Zhang
202003151126@zjut.edu.cn

Syauqie Muhammad Marier
syauqie.mm@gmail.com

Yichen Liu
liuien@outlook.com

Yuhao Chen
yuhaochen4859@gmail.com

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

2   School of Computing Technologies, RMIT University, Melbourne 3000, Australia

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 [113, 191]. It finds extensive applications in various domains closely intertwined with our lives, including energy production and consumption [55, 143, 194, 209, 240, 290], meteorological variations [177, 275], finance, stock markets, and econometrics [6, 29, 37, 119, 163, 213, 224] sales and demand [17, 23] urban traffic flows [144, 165, 237], and welfare-related healthcare conditions [120, 173, 192, 238].

Machine learning, data science, and other research groups employing operations research and statistical methods have extensively explored time series forecasting [71–73, 81]. 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 [203] and rapid development of deep learning [94, 134], 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.

In recent years, deep learning methods have continuously advanced and innovated in time series forecasting (TSF) across various domains [18, 132, 150, 183, 196, 206, 257]. 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:

- *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.
- *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.
- *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.

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 Sect. 6. In Appendix A, an exhaustive account of TSF datasets across various domains is presented. Figure 1 shows an outline of the entire paper.

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 [101]. It is commonly employed for the analysis and prediction of trends and patterns [175] that evolve over time.

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.

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 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

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 $\mathbb{Z} = (0, \pm 1, \pm 2, \ldots)$ represents the set of positive and negative integers [81]. When only a limited amount of data is available, a time series can be represented as $(y_1, y_2, y_3, \ldots)$. 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 $[t_1, t_2]$.

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
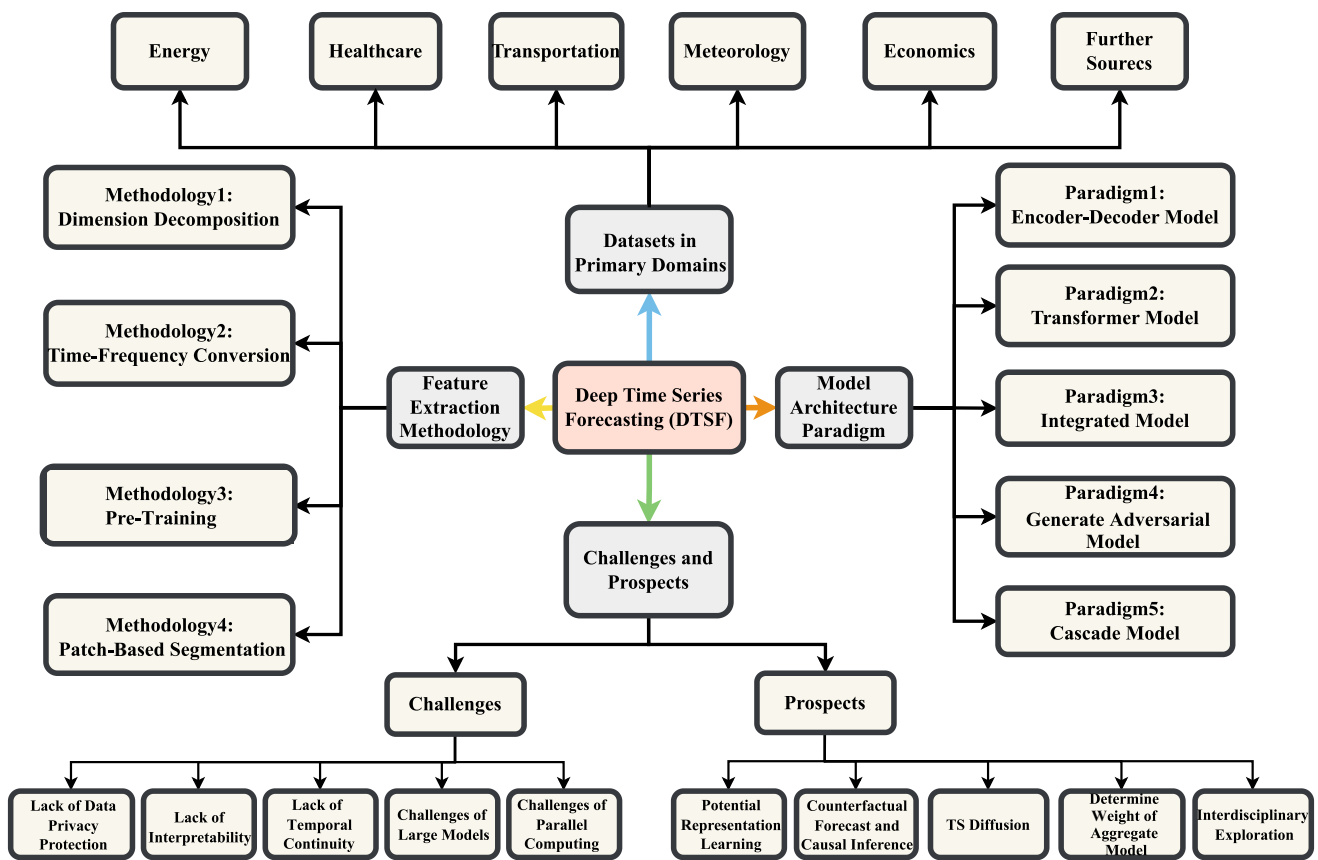
Fig. 1 The outline of this article

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 [276]. 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 [262]. 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 [189], 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.

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 [204], 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 [193]. Additionally, seasonal adjustment, periodic modeling, and the integration of road network knowledge are effective methods for addressing external influences. For instance, MultiSPANS [299] uses a structural entropy minimization algorithm to generate optimal road network hierarchies, considering complex multi-distance dependencies in the road network for prediction; [128], in summarizing forecasting tasks, constructed a new bus station distance network to account for the relationships between external bus stations.

On the other hand, in addition to being categorized as Univariate [115, 174, 211, 281] and Multivariate [125, 164] 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.

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.

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

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) [24, 50, 109] and simple Exponential Smoothing (ES) [87], were based on time series.

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

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

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

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) [24, 109, 135] 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 [1, 24, 109] (ARMA, as shown in equation 5). These models aim to accurately capture the auto correlation and averaging properties of time series data.

$$AR : Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \xi_t \qquad (3)$$

$$MA : Y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} \qquad (4)$$

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

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.

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 [24, 50, 102, 109, 280] 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:

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

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.

Machine learning models represented by Random Forests and Decision Trees [5, 111, 129, 201] offer enhanced flexibility and predictive performance in statistical forecasting [2, 105]. 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.

However, the development of emerging technologies such as the Internet of Things (IoT) has brought efficiency and convenience to data acquisition, collection, and storage [127, 140]. The era of big data has arrived [74, 205], 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 [34, 186, 255]. Different industries and domains are also increasingly in need of accurate forecasting models to support decision-making and planning [200]. Furthermore, more complex relationships among data are encountered in practical applications, requiring more flexible and accurate models to tackle these challenges.

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

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 [250, 251, 284, 294], providing valuable insights for subsequent research.

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 [206] leveraged RNN and autoregressive techniques to capture temporal dependencies and patterns in time series data. MQRNN [248] 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 [225] proposed a dynamic computational graph neural network with a standard exponential smoothing model and LSTM in a common framework.

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.

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 [15]. 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 [56]. In traffic flow prediction, TCNs have been successfully applied to model

the temporal dependencies in sensor data, achieving high accuracy in forecasting traffic conditions [289]. 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.

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.

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 [202] and various forms of Transformers are well-known applications.

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 [231]. 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 [46]. 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 [15]. 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 [45]. 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.

Encoder-decoder has also been extensively and successfully applied in the field of TSF. For instance, [190] was inspired by U-net [202] 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.

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 [251, 265, 292, 294]. 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.

Likewise, Guo et al. [97] 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.

There are some other examples of encoder-decoder here as well. In [21], 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 (see Fig. 3). There are also network architectures that explicitly involve an encoder but lack an explicit decoder [66].

Table 1 DTSF model architecture paradigm

Architecture	Model	Multi/uni	Output	Loss	Metrics	Year
	COST [249]	Multi & uni	Point	contrastive loss	MSE, MAE	2022
	TS2Vec [274]	Multi & uni	Point	contrastive loss	MSE	2022
	ACT [146]	Multi & uni	Point	cross-entropy	Q50 loss, Q90 loss	2022
	SimTS [291]	Multi & uni	Point	cos-similarity loss, InfoNCE loss	MAE, MSE	2023
	DeepTCN [41]	Multi	Pro	quantile loss	NRMSE, SMAPE, MASE	2020
	STEP [215]	Multi	Pro	MAE	MAE, RMSE, MAPE	2022
	DCAN [106]	Multi	Point	RMSE	MAE, RMSE	2022
	FusFormer [265]	Multi	Point	–	RMSE, RMSE Decrease	2022
	HANet [21]	Multi	Point	–	MAE, RMSE	2022
	D^3VAE [145]	Multi	Pro	–	MSE, CRPS	2022
	TI-MAE [147]	Multi	Point	MSE	MSE, MAE	2023
Encoder - Decoder	AST [254]	Uni	Pro	cross-entropy	Q50, Q90 loss	2020
	TFT [150]	Multi & uni	Prob	quantile loss	P50, P90 quantile loss	2021
	Informer [292]	Multi & uni	Point	MSELoss	MSE, MAE	2021
	ETSformer [250]	Multi & uni	Point	MSELoss	MSE, MAE	2022
	FEDformer [294]	Multi & uni	Point	MSELoss	MSE, MAE, Permutation	2022
	TACTiS [62]	Multi & uni	Pro	log-likelihood	CRPS-Sum, CRPS-means	2022
	Autoformer [251]	Multi & uni	Point	L2 loss	MSE, MAE	2022
	NSTformer [159]	Multi & uni	Point	L2 loss	MSE, MAE	2023
	Dateformer [270]	Multi & uni	Point	MSE	MSE, MAE	2023
	Crossformer [286]	Multi & uni	Point	MSE	MSE, MAE	2023
	Scaleformer [214]	Multi & uni	Pro	MSE	MSE, MAE	2023
	BasisFormer [179]	Multi & uni	Point	MSE	MSE, MAE	2023
	CRT [282]	Multi	Point	–	ROC-AUC, F1-Score	2021
	Pyraformer [156]	Multi	Point	MSE	MSE, MAE	2022
	TDformer [284]	Multi	Point	MSE	MSE, MAE	2022
	FusFormer [265]	Multi	Point	–	RMSE, RMSE Decrease	2022
	Scaleformer [214]	Multi	Point	MSE, Huber, Adaptive loss	MSE, MAE	2022
	Infomaxformer [234]	Multi	Pro	MSELoss	MSE, MAE	2023
	PatchTST [180]	Multi	Point	Adaptive Loss	MSE, MAE	2023
Transformer	iTransformer [160]	Multi	Point	L2 Loss	MSE, MAE	2023
	MCformer [103]	Multi	Point	MSE, MAE	MSE, MAE	2024
	SAMformer [114]	Multi	Point	MSE	MSE, MAE	2024
	TSLANet [68]	Multi	Point	MSE	MSE, MAE	2024
	MASTER [142]	Multi	Point	MSE	IC, ICIR, RankIC	2024
	TimeSiam [60]	Multi	Point	L2, Cross-Entropy	MSE, MAE, Recall, F1 Score	2024
	Chronos [12]	Multi	Point	Cross Entropy	WQL, CRPS, MASE	2024
	TimeXer [246]	Multi	Point	L2 loss	MSE, MAE	2024
	Time-SSM [112]	Multi	Point	MSE	MSE, MAE	2024
	SageFormer [288]	Multi	Point	MSE	MSE, MAE	2024
	TIME-LLM [244]	Multi	Point	MSE, SMAPE	MSE, MAE, SMAPE	2024
	CARD [245]	Multi	Point	MSE, MAE	MSE, MAE	2024
	Pathformer [38]	Uni	Pro	L1 loss	MSE, MAE	2024
	ForGAN [130]	Multi & uni	Pro	RMSE	MAE, MAPE, RMSE	2019
	COSCI-GAN [212]	Multi & uni	Pro	Global loss = local + central	MAE	2022
	RCGAN [70]	Multi	Pro	cross-entropy	AUROC, AUPRC	2017
	TimeGAN [269]	Multi	Pro	Unsupervised, Supervised, Reconstruction, Loss	Discriminative and Predictive Score	2019
	PSA-GAN [116]	Multi	Point	Wasserstein loss	–	2022
	AEC-GAN [243]	Multi	Point	MSE	ACF, Skew / Kurt, FD	2023

Table 1 (continued)

Architecture	Model	Multi/uni	Output	Loss	Metrics	Year
	ITF-GAN [121]	Multi	Point	MSE	MSE, STS, Pearson, Hellinger, Pred	2024
	MAGAN [76]	Multi	Point	–	MAE, MAPE	2024
	TSGAN [258]	Multi	Point	–	MAE, RMSE, MAPE	2022
GAN	AST [254]	Uni	Pro	cross-entropy	Q50 loss, Q90 loss	2020
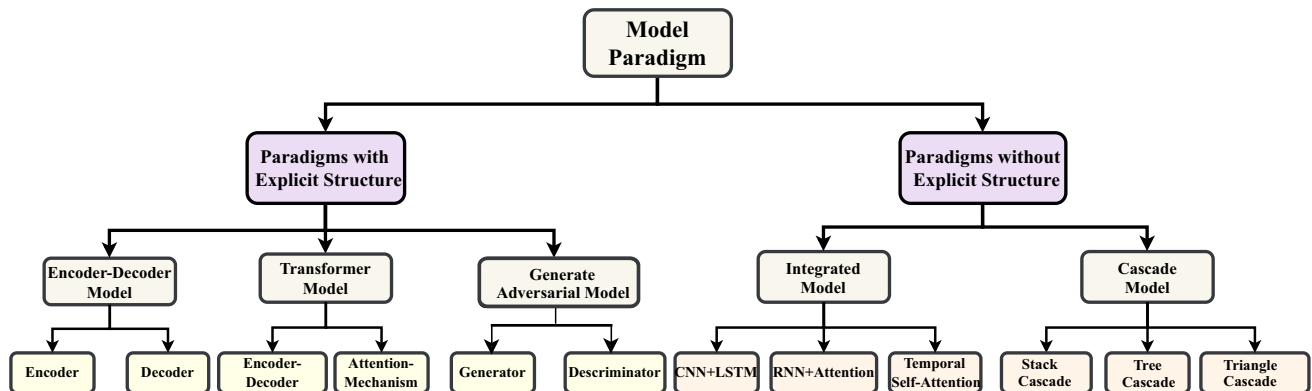
Fig. 2 The details of five paradigms

Table 2 A comparative analysis of time series forecasting models: advantages, disadvantages, applications, and prediction lengths

Model	Advantages	Disadvantages	Applications	Prediction horizon
Informer [292]	Effcient; strong repre–sentation; good generalization	Sensitive to data shifts	Energy; weather	Long
HANet [21]	Capture complex dependencies; flexible for multivariate data	High complexity	Weather; ecology	Long
Autoformer [251]	Efficient; good information	High complexity; depend on data periodicity	Finance; energy; electricity; traffic; weather; healthcare	Long
ETSformer [250]	Combine traditional methods with transformer; adaptive time window	High computational cost; requires large data	Finance; energy; electricity; traffic; weather; healthcare	Short
FEDformer [294]	Frequency enhancement; Better flexibility for long-term forecasts	High complexity; large data needed	Finance; energy; electricity; traffic; weather; healthcare	Long
TreeDRNet [295]	Capture time dynamics; efficient training with joint networks	High complexity; need large data	Finance; energy; electricity; traffic; weather; healthcare	Long
TATCN [242]	Capture temporal dependencies; extract local patterns	High computational cost; data dependence	Electricity; healthcare	Short

3.1.2 Transformer model

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 complexity, high memory usage, and inherent limitations of the encoder-decoder architecture. To address these limitations, Informer [292] was an efficient Transformer-based architecture specifically designed for
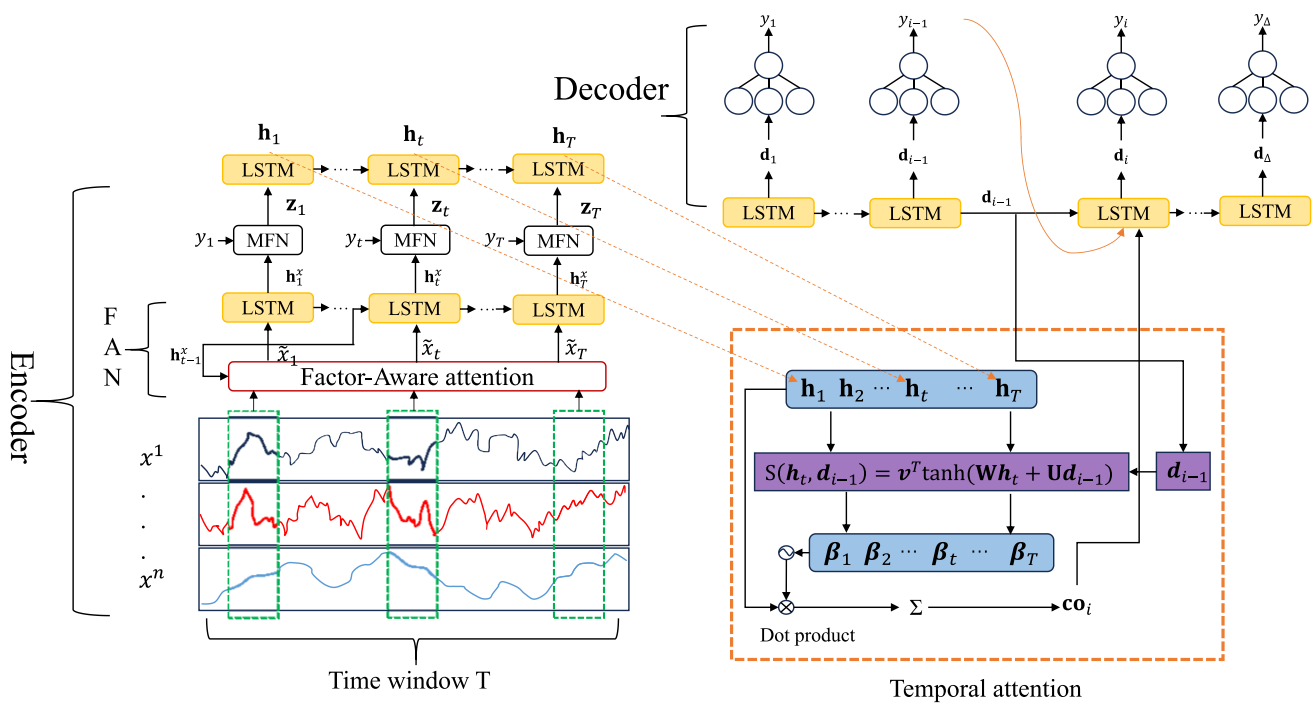
Fig. 3 The overview of HANet model

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 [292] 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 [150] 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 (see Fig. 4).

Autoformer [251], on the other hand, argues that previous Transformer-based prediction models (e.g., Informer [292]) mainly focused on improving self-attention for sparse

Fig. 4 The overview of Informer model
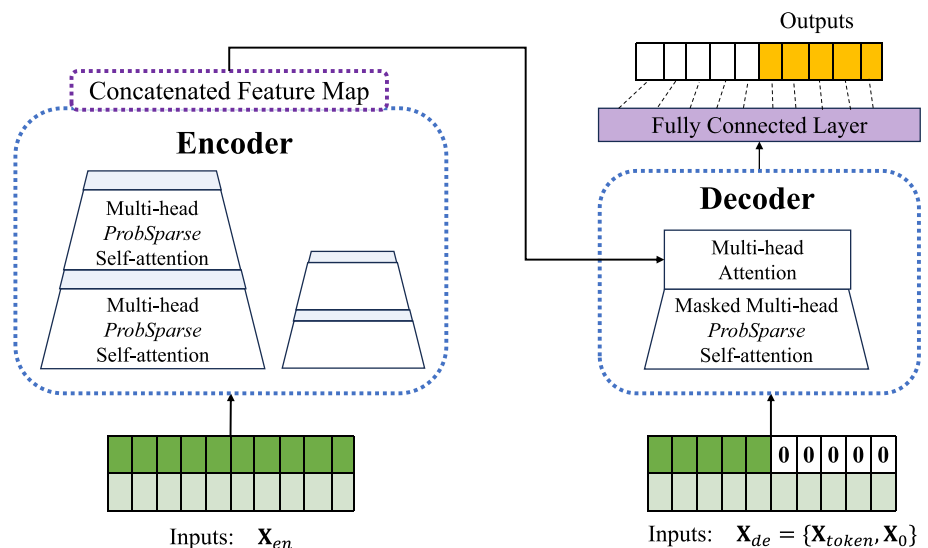
versions. While significant performance improvements were achieved, they sacrificed the utilization of information. One of the reasons why Transformer cannot be directly applied to LSTF is the complex characteristics of time series data. Without special design, traditional attention mechanisms struggle to model and learn these characteristics. Autoformer [251] adopts decomposition as a standard approach for time series analysis [49, 167], as it is believed that decomposition can untangle the intertwined time patterns and highlight the intrinsic properties of time series. Autoformer [251] introduces a novel decomposition architecture with autocorrelation mechanisms, 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 (Fig. 5).

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 reducing computational burden. TSMixer [65] 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) [285] 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 [250] 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.

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.

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 [19].
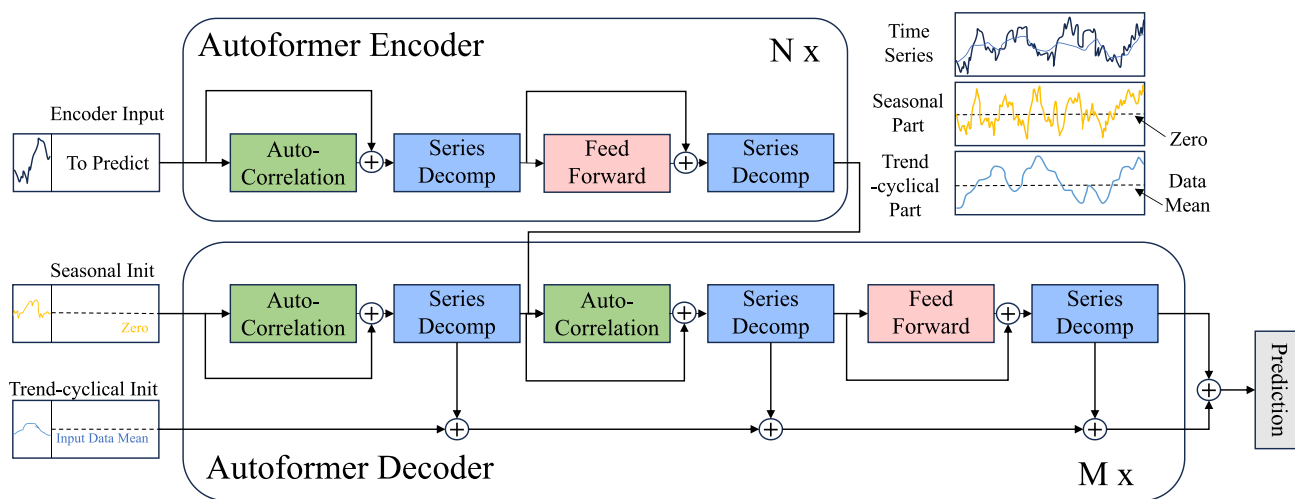

Fig. 5 The overview of autoformer model

In the existing literature on GAN-based TSF, most studies focus on generating synthetic time series datasets [70, 233, 269]. 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 [269], 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 [233]. The study introduced two variations, Recursive GAN (RGAN) and Recursive Conditional GAN (RCGAN), where RGAN generates real-valued data sequences, and RCGAN generates sequences conditioned on specific inputs. The discriminators and generators of both RGAN and RCGAN are based on simple RNN architectures.

Furthermore, a deep neural network-based approach was proposed for modeling financial time series data [233]. 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 (see Fig. 6).

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" [108]. 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 [15], 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 (see Fig. 7).

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 [14, 15, 218]. In [218], precipitation forecasting was modeled

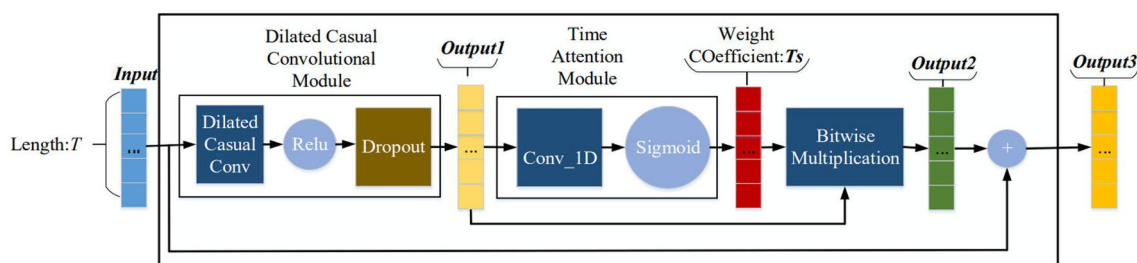**Fig. 6** The overview of TimeGAN model
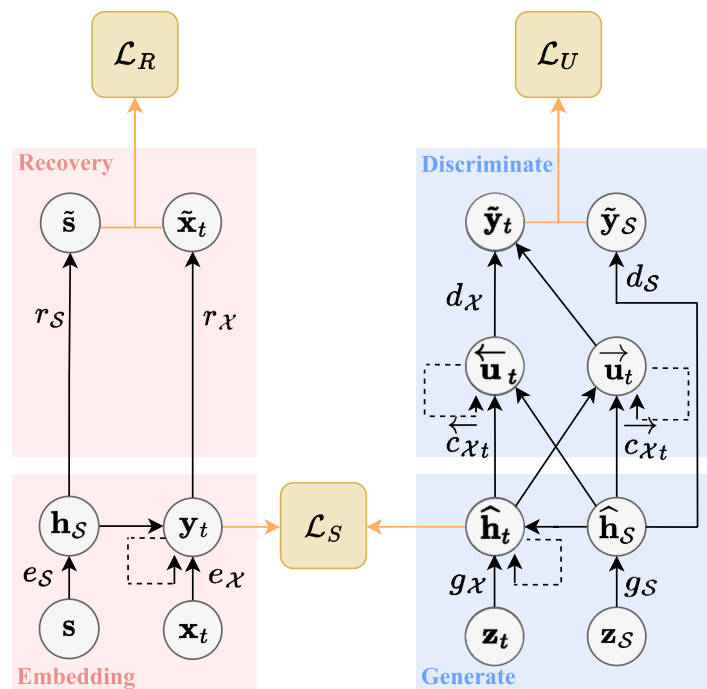
Fig. 7 The overview of TATCN model

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) [14] 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.

3.2.2 Cascade model

Cascade networks, which are widely used in deep neural networks, especially in Computer Vision (CV) domain [28], 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 (Fig. 8).

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 considered as utilizing the cascade idea, as seen in the Transformer series [49, 54, 141, 167]. Additionally, some models [295] 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

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.

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
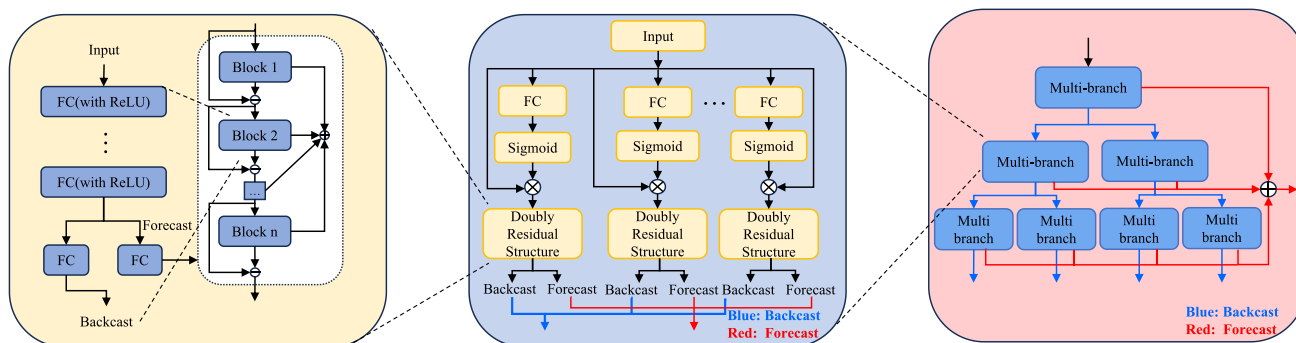

Fig. 8 The overview of TreeDRNet model

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

In general, time series data can be decomposed into three main components: trend, seasonality, and residuals or white noise [219], as illustrated in Fig. 9.

4.1.1 Trend

Represents the long-term changes in the time series data and reflects the overall growth or decline of the data over an extended period [175]. For example, the increase in population over the years exhibits an upward trend [1], 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.

4.1.3 Residuals

Represent the part of the data that cannot be explained by the trend and seasonality components [170]. 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.

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.

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 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.

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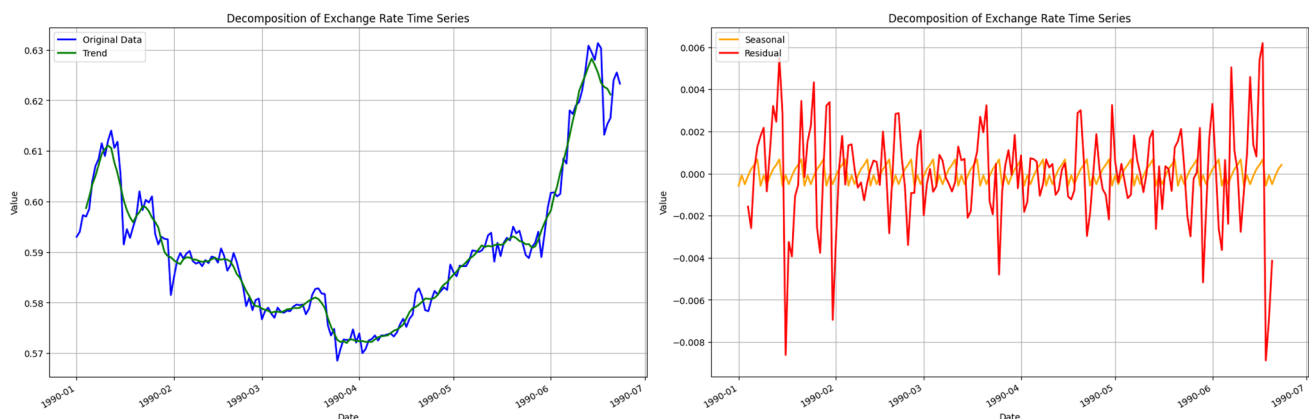
Fig. 9 Components of the time series. The data is sourced from the Exchange-Rate dataset spanning from January 1, 1990, to June 23, 1990. The blue line represents the original data, the green indicates the trend, the yellow represents seasonality, and the red signifies the residuals

4.2.1 Dimension decomposition

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 (Fig. 10).

In current research, some works have integrated encoder-decoder architectures with seasonal-trend decomposition [32, 188, 234, 247, 251, 284, 294, 298]. Wu et al. [251] in the similar work, devised an internal decomposition block to endow deep forecasting model with intrinsic progressive decomposition capability. Subsequently, Zhou et al. [294] proposed a seasonal-trend-based frequency enhanced decomposition Transformer architecture in the FEDformer framework. Additionally, Wang et al. [247] introduced the LaTS model, leveraging variational inference to unravel latent space seasonal trend features, and Zhang et al. [284] presented the TDformer model, using MLP to model trends and Fourier attention to simulating seasonality. Notably, Zhu et al. [298] designed an approach to decompose input sequences into trend and residual components across multiple scales, 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 [10]. 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.

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 (Fig. 11).

In current research, the time-frequency domain conversion finds extensive application in the preprocessing and feature extraction of time series data [43, 131, 229]. 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 [294]. 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 the primary fluctuation patterns across frequencies. In their work, [30] 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. [266] 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
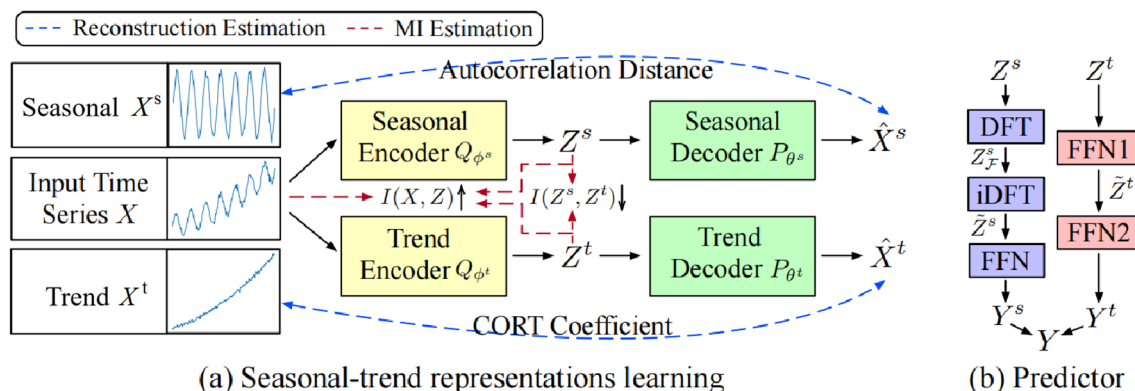

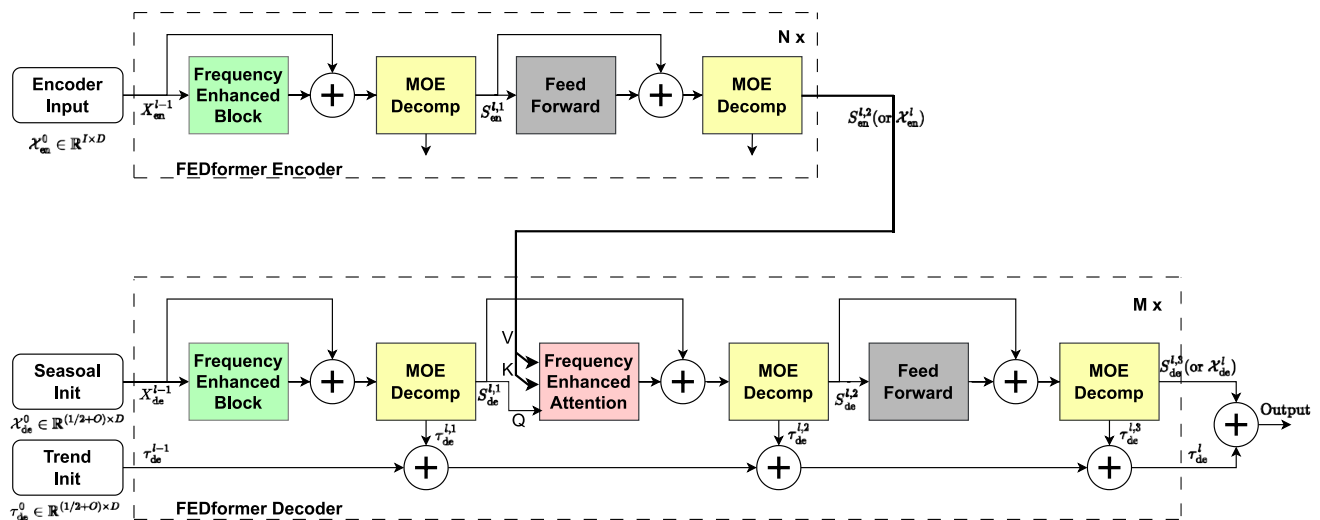**Fig. 10** The overview of LaST model

Fig. 11 The overview of FEDformer model

channel and temporal correlations across both inter-series and intra-series scales.

Furthermore, employing time-frequency domain conversion can help reduce the impact of noise and interference [96, 293]. 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. [293] 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.

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

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 [44, 198, 207, 230]. In certain scenarios, the adoption of sampling pre-training methods could be considered (Fig. 12).

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 [20, 222], 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 [184]. In time series, if the representations based on time and frequency for the same instance are close in the time-frequency space, it suggests a certain similarity or consistency in their features or attributes. Zhang et al. [282]. 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 (Fig. 13).

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
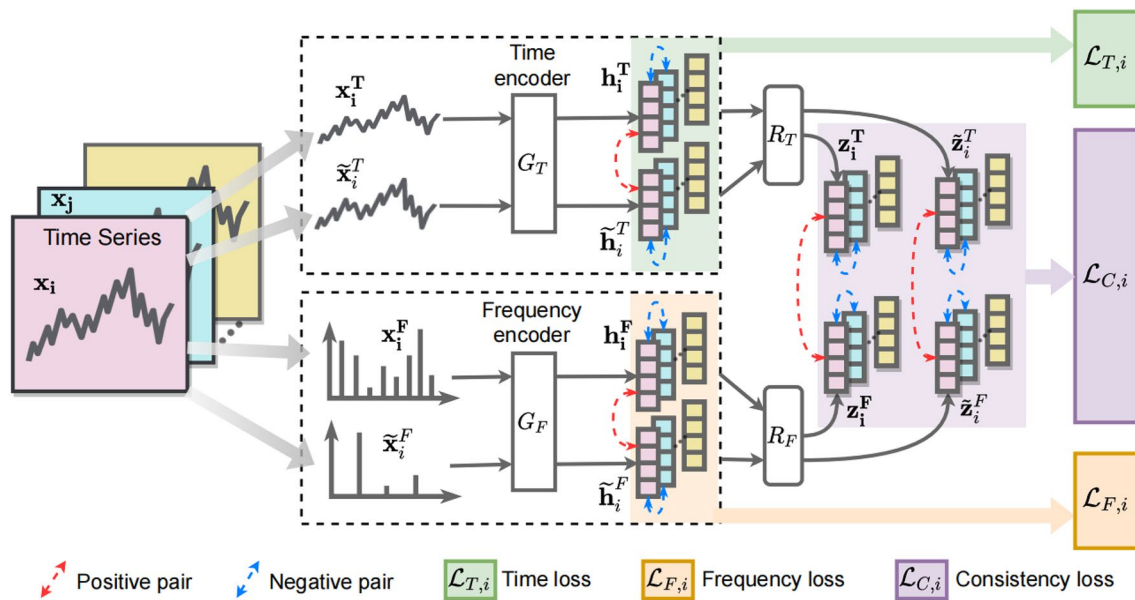
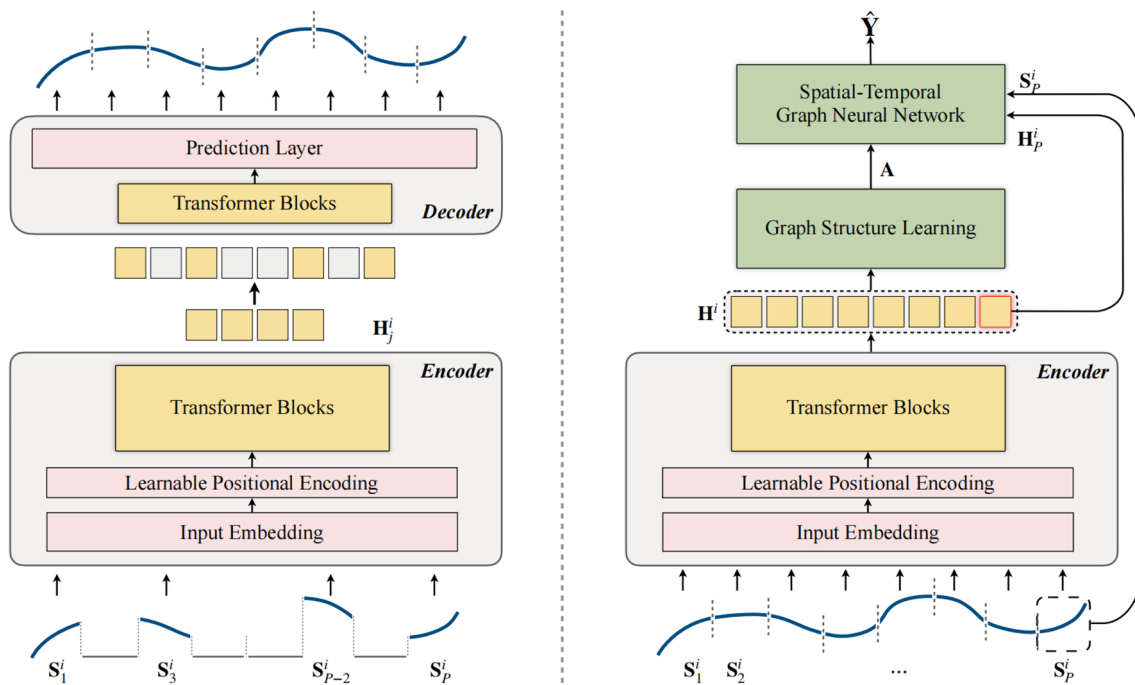**Fig. 12** The overview of TF-C

Fig. 13 The overview of STEP

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.

In the work STEP, Shao et al. [215] 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 [147] exhibits analogous efficacy in this regard. In the pre-training model SimMTM, Dong et al. [59] 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.

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

4.2.4 Patch-based segmentation

In recent DTSF works, especially those of the Transformer models, the adoption of patch-based data organization has become prevalent [65, 93, 151, 180, 261, 285]. 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 (Fig. 14).

Nie et al. [180] 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 [287]. It allows for spatial aggregation of representations for global tokens in the graph.

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.

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.
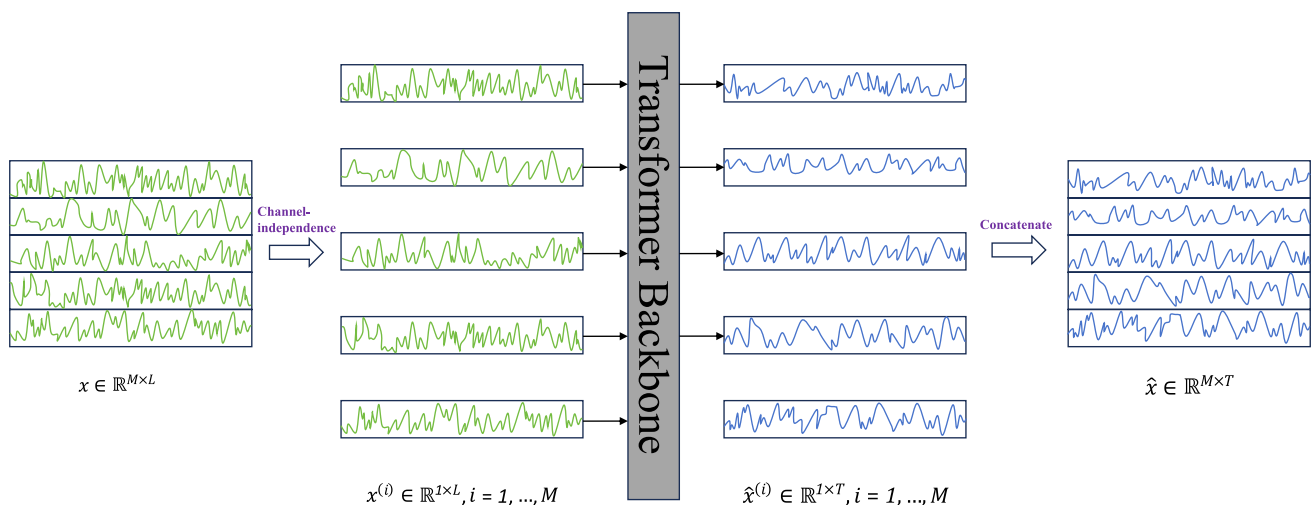

Fig. 14 The overview of PatchTST

5.1 Challenges

5.1.1 Lack of data privacy protection and completeness

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 [171]. 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 [25, 83, 232], traffic speed and flow [158, 279], energy consumption [208, 283], and communication networks [58, 227], 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 [176], 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 [185]. This enables users to better comprehend the model's predictive outcomes while maintaining high forecasting precision.

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 [161, 199]. The other category encompasses gradient-based methods, such as those presented in [124, 220, 221], which scrutinize the network gradients to determine which input features exert the most significant influence on the loss function.

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 [16, 47, 141]. 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 [150], such as seasonal patterns.

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 [8, 69]. Among these, sparse optimization methods, which obviate the need for time-consuming back-propagation training, exhibit efficient training capabilities on CPUs. These methods offer insights for further exploration into interpretable time series forecasting.

5.1.3 Lack of temporal continuity

Compared to traditional deep learning forecasting models, the proposal of the Neural Ordinary Differential Equation (NODE) [39] has directed our attention 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 [53, 85, 138], handling irregular data [210], achieving interpretable continuity [84, 117], optimizing model parameters [42], and exploring differential dynamics [90, 153]. The ETN-ODE model proposed by Gao et al. [84] 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 [85] 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.

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

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

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 [259]. 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 [33, 118, 296], inference of medical data [95, 228], forecasting human mobility trajectories [31], and serving as general-purpose models for weather and energy demand predictions [137, 157, 256, 273, 278].

On another note, significant strides have been made in the training of foundational time series models [88, 260]. The recent TimeGPT-1 model [197] 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.

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.

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 [51, 67]. Some works employ multi-module architectures or model ensembles [166, 172, 264], while others use pre-training with denoising, smoothing properties, siamese structures or 2D-variation modeling [252, 277, 291], 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 [162, 187, 274, 282] 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.

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.

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 [3, 104, 268], 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 [22, 86, 139, 149, 154].

Moreover, while some efforts provide counterfactual explanations for time series models [57, 178], 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 [263].

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

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 [196], while CSDI utilizes non-autoregressive methods with self-supervised masking [235]. Similarly, SSSD utilizes structured state-space models to reduce computational complexity [4]. Despite being early explorations in the TSF domain, these models still suffer from slow inference, high complexity, and boundary inconsistencies.

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 [126]. TimeDiff addresses boundary inconsistencies with future mixups and autoregressive initialization mechanisms [216]. The multi-scale diffusion model MR-Diff leverages multi-resolution temporal structures for sequential trend extraction and non-autoregressive denoising [9].

The first framework based on DDPM, Diffusion-TS, accurately reconstructs samples using Fourier-based loss functions, extending to forecasting tasks [7]. Furthermore, the TMDM model combines conditional diffusion generation processes with Transformer to achieve precise distribution prediction for multivariate time series [11].

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 [13, 169, 236]. 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. [79] 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 [148], 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 [80, 100] presents a unique utility in determining ensemble model weights and is a path worth exploring.

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.

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.

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

Table 3 Time series datasets in primary domains

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

The table summarizes commonly used datasets and indicates whether they are multivariate, which implies temporal alignment with known timestamps

challenges in the field and propose some potential research directions.

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 [89], 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 [99]. Time alignment techniques, such as the TAM model, synchronize data from different time granularities by introducing a novel time-distance measure [77]. Deep generative models, like GinAR, address missing values and noise by generating new samples and rebuilding spatiotemporal dependencies [272]. Self-supervised learning methods, such as SimCLR, allow models to learn from unlabeled data, improving adaptability to heterogeneous sources [40]. Finally, collaborative attention mechanisms capture complex correlations between multimodal data and adjust modality weights dynamically, enhancing model learning capacity [61]. These models and techniques effectively integrate heterogeneous data, improving the stability and accuracy of time series forecasting in multi-source environments.
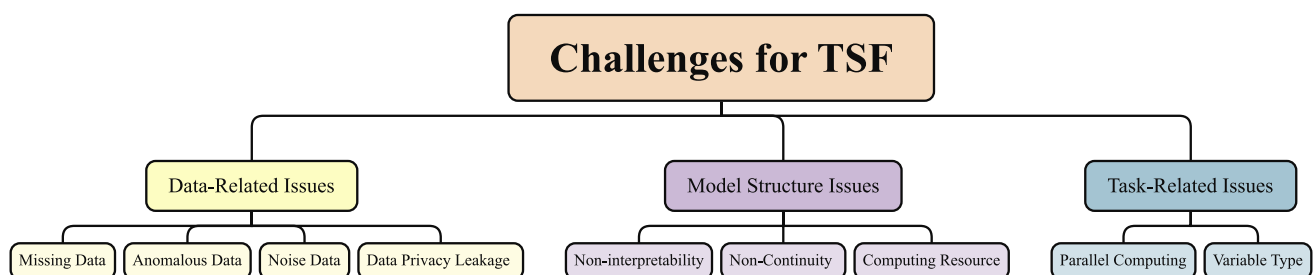
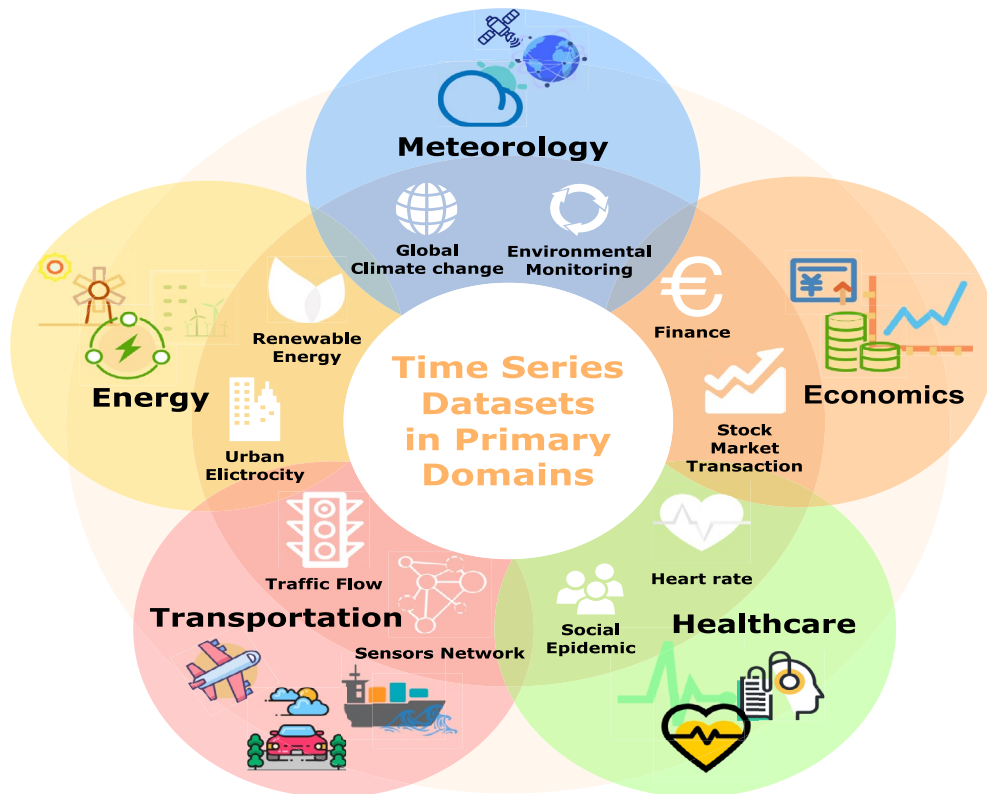**Fig. 15** Challenges in time series forecasting

Fig. 16 Time series datasets in primary domains

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 [223], wind energy [75], and solar energy [194]. 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 (see Fig. 16).

Electricity transformer temperature (ETT)

The ETT-small dataset encompasses data originating from two distinct power transformer installations, each situated at a separate site [292]. 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 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 h 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.

Electricity

The initial dataset utilized in this investigation is the Electricity Load Diagrams 2011-2014 Dataset [241], which records 370 customers' electricity usage information between 2011 and 2014. Data is recorded in the original dataset every 15 min. 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.

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.

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]

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.

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.

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.

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

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.

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.

PeMSD4/7/8

These datasets are highly regarded as industry standards for traffic forecasting [35].

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.

---

[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/

[5] https://pems.dot.ca.gov/

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.

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.

Weather2

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

Temperature rain

Consisting of 32,072 daily time series, this dataset [91] 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.

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.

Exchange-rate

This dataset [132] 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.

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 [181], 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 [182]+"+. 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.

Dominick

This dataset [92] 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.

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.

---

[6] https://www.ncei.noaa.gov/data/local-climatological-data/

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

Table 4 Summary of the datasets used in the experiments

Domain	Variants	Dataset	Data time range	Data granularity	References
	21	the Scada wind farm in Turkey	2018/1/1-2018/12/29	10 m	[152]
	–	Global horizontal solar radiation data	1998/1/1-2007/12/1	1 h	[226]
Energy	–	Rooftop PV plant	2015/1/1-2016/12/31	30 m	[239]
	9	UCI household electric power consumption	2006/12-2010/11	1 m	[26]
	–	Spanish electricity demand	2014/01/02-2019/11/01	10 m	[133]
	–	Electric vehicles power consumption	2015/3/2-2016/5/31	1 h	[133]
	–	CDC ILI data	2010-2018	1 d	[253]
	45	DEAP	-	1 interval	[123]
Healthcare	9	Turkish COVID-19 data	2020/3/27-2020/6/11	1 d	[122]
	9	COVID-19 dataset of Orissa state	2020/1/30-2020/6/11	1 d	[52]
	207	METR-LA	2012/3/1-2012/6/30	5 m	[27]
Transportation	325	PeMS-BAY	2017/1/1-2017/5/31	5 m	[27]
	–	BJER4	2014/7/1-2014/8/31	5 m	[271]
	6	Daily data of Shenzhen	from 2015	-	[36]
Meteorology	–	CHIRPS	1981-2015	-	[82]
	–	WeatherBench	-	-	[195]
	5	S&P500	1997/1/1-2016/12/1	1 d	[136]
Economics	13	NSE stocks data	1996/1/1-2015/6/30	1 d	[110]
	6	NYSE stock data	2011/1/3-2016/12/30	1 m	[110]

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.

M4

The M4 dataset [168] encompasses 100,000 real-life series in diverse domains, including micro, industry, macro, finance, demographic, and others.

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.

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.

Data Availability Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Declarations

Conflict 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.

---

[8] https://forecasters.org/resources/time-series-data/

[9] https://mofc.unic.ac.cy/m5-competition/

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

References

1. Adhikari Ratnadip, Agrawal Ramesh K (2013) An introductory study on time series modeling and forecasting. *arXiv preprint*[SPACE]arXiv:1302.6613. Accessed 2 Feb 2025

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

3. Alaa Ahmed M, Weisz Michael, Van Der Schaar Mihaela (2017) Deep counterfactual networks with propensity-dropout. *arXiv preprint*[SPACE]arXiv:1706.05966. Accessed 2 Feb 2025

4. Lopez Alcaraz JM, Strodthoff N (2022) Diffusion-based time series imputation and forecasting with structured state space models. Transactions on Machine Learning Research. arXiv:2208.09399. Accessed 3 Feb 2025

5. Ali Jehad, Khan Rehanullah, Ahmad Nasir, Maqsood Imran (2012) Random forests and decision trees. Int J Comput Sci Issues (IJCSI) 9(5):272

6. Andersen, Torben G., Bollerslev, Tim, Christoffersen, Peter, Diebold Francis X (2005) Volatility forecasting

7. Yuan X, Qiao Y (2024) Diffusion-TS: Interpretable diffusion for general time series generation. In: The Twelfth International Conference on Learning Representations. arXiv:2403.01742. Accessed 3 Feb 2025

8. Liu X, Chen D, Wei W, Zhu X, Yu W (2024) Interpretable sparse system identification: Beyond recent deep learning techniques on time-series prediction. In: The Twelfth International Conference on Learning Representations

9. Shen L, Chen W, Kwok J (2024) Multi-resolution diffusion models for time series forecasting. In: The Twelfth International Conference on Learning Representations. https://openreview.net/forum?id=mmjnr0G8ZY. Accessed 3 Feb 2025

10. Park J, Gwak D, Choo J, Choi E (2024) Self-supervised contrastive forecasting. In: The Twelfth International Conference on Learning Representations. arXiv:2402.02023. Accessed 3 Feb 2025

11. Li Y, Chen W, Hu X, Chen B, Zhou M (2024) Transformer-modulated diffusion models for probabilistic multivariate time series forecasting. In: The Twelfth International Conference on Learning Representations. https://openreview.net/forum?id=qae04YACHs. Accessed 3 Feb 2025

12. Ansari AF, Stella L, Turkmen C, Zhang X, Mercado P, Shen H, Shchur O, Rangapuram SS, Arango SP, Kapoor S, Zschiegner J (2024) Chronos: learning the language of time series. Transactions on Machine Learning Research. arXiv:2403.07815. Accessed 3 Feb 2025

13. Arbib Michael A (2003) The handbook of brain theory and neural networks. MIT Press, New York

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

15. Bai Shaojie, Kolter J Zico, Koltun Vladlen (2018) An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint*[SPACE]arXiv:1803.01271. Accessed 2 Feb 2025

16. Bai Tian, Zhang Shanshan, Egleston Brian L, Vucetic Slobodan (2018) 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. pp 43–51

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

18. Bandara Kasun, Bergmeir Christoph, Hewamalage Hansika (2020) Lstm-msnet: leveraging forecasts on sets of related time series with multiple seasonal patterns. IEEE Trans Neural Netw Learn Syst 32(4):1586–1599

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

20. Berthelot David, Roelofs Rebecca, Sohn Kihyuk, Carlini Nicholas, Kurakin Alexey (2022) Adamatch: A unified approach to semi-supervised learning and domain adaptation. In *International Conference on Learning Representations*. URL https://openreview.net/forum?id=Q5uh1Nvv5dm. Accessed 2 Feb 2025

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

22. Bica I, Alaa AM, Jordon J, van der Schaar M (2020) Estimating counterfactual treatment outcomes over time through adversarially balanced representations. In: International Conference on Learning Representations (ICLR). arXiv:2002.04083. Accessed 3 Feb 2025

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

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

25. Briggs Christopher, Fan Zhong, Andras Peter (2022) Federated learning for short-term residential load forecasting. IEEE Open Access J Power Energy 9:573–583

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

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

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

29. Callot Laurent AF, Kock Anders B, Medeiros Marcelo C (2017) Modeling and forecasting large realized covariance matrices and portfolio choice. J Appl Economet 32(1):140–158

30. Cao Defu, Wang Yujing, Duan Juanyong, Zhang Ce, Zhu Xia, Huang Congrui, Tong Yunhai, Bixiong Xu, Bai Jing, Tong Jie et al (2020) Spectral temporal graph neural network for multivariate time-series forecasting. Adv Neural Inf Process Syst 33:17766–17778

31. Cao Defu, Jia Furong, Arik Sercan O, Pfister Tomas, Zheng Yixiang, Ye Wen, Liu Yan (2023a) Tempo: Prompt-based generative pre-trained transformer for time series forecasting. *arXiv preprint* [SPACE]arXiv:2310.04948. Accessed 2 Feb 2025

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

33. Chang Ching, Peng Wen-Chih, Chen Tien-Fu (2023) Llm4ts: Two-stage fine-tuning for time-series forecasting with pre-trained llms. *arXiv preprint* [SPACE]arXiv:2308.08469. Accessed 2 Feb 2025

34. Che Dunren, Safran Mejdl, Peng Zhiyong (2013) From big data to big data mining: challenges, issues, 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*. Springer, New York. pp 1–15

35. Chen Chao, Petty Karl, Skabardonis Alexander, Varaiya Pravin, Jia Zhanfeng (2001) Freeway performance measurement system: mining loop detector data. Transp Res Rec 1748(1):96–102

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

37. Chen Mu-Yen, Chen Bo-Tsuen (2015) A hybrid fuzzy time series model based on granular computing for stock price forecasting. Inf Sci 294:227–241

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

39. Chen Ricky TQ, Rubanova Yulia, Bettencourt Jesse, Duvenaud David K (2018) Neural ordinary differential equations. Adv Neural Inf Process Syst. p 31

40. Chen Ting, Kornblith Simon, Norouzi Mohammad, Hinton Geoffrey (2020) A simple framework for contrastive learning of visual representations. In: International Conference on Machine Learning. PMLR pp 1597–1607

41. Chen Yitian, Kang Yanfei, Chen Yixiong, Wang Zizhuo (2020) Probabilistic forecasting with temporal convolutional neural network. Neurocomputing 399:491–501

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

43. Chen Yushu, Liu Shengzhuo, Yang Jinzhe, Jing Hao, Zhao Wenlai, Yang Guangwen (2023) A joint time-frequency domain transformer for multivariate time series forecasting. *arXiv preprint* [SPACE]arXiv:2305.14649. Accessed 2 Feb 2025

44. Cheng Joseph Y, Goh Hanlin, Dogrusoz Kaan, Tuzel Oncel, Azemi Erdrin (2020) Subject-aware contrastive learning for biosignals. *arXiv preprint* [SPACE]arXiv:2007.04871. Accessed 2 Feb 2025

45. Cheng Qi, Chen Yixin, Xiao Yuteng, Yin Hongsheng, Liu Weidong (2022) A dual-stage attention-based bi-lstm network for multivariate time series prediction. J Supercomput 78(14):16214–16235

46. Cho Kyunghyun (2014) Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint* [SPACE]arXiv:1406.1078. Accessed 2 Feb 2025

47. Choi Edward, Bahadori Mohammad Taha, Sun Jimeng, Kulas Joshua, Schuetz Andy, Stewart Walter (2016) Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. Adv Neural Inf Process Syst. p 29

48. Cirstea Razvan-Gabriel, Guo Chenjuan, Yang Bin, Kieu Tung, Dong Xuanyi, Pan Shirui (2022) Triformer: Triangular, variable-specific attentions for long sequence multivariate time series forecasting–full version. *arXiv preprint* [SPACE]arXiv:2204.13767. Accessed 2 Feb 2025

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

50. Cochrane John H (1997) Time series for macroeconomics and finance

51. Darban Zahra Zamanzadeh, Webb Geoffrey I, Pan Shirui, Salehi Mahsa (2023) Carla: A self-supervised contrastive representation learning approach for time series anomaly detection. *arXiv preprint* [SPACE]arXiv:2308.09296. Accessed 2 Feb 2025

52. Dash Satyabrata, Chakravarty Sujata, Mohanty Sachi Nandan, Pattanaik Chinmaya Ranjan, Jain Sarika (2021) A deep learning method to forecast covid-19 outbreak. N Gener Comput 39(3–4):515–539

53. De Brouwer Edward, Simm Jaak, Arany Adam, Moreau Yves (2019) Gru-ode-bayes: continuous modeling of sporadically-observed time series. Adv Neural Inf Process Syst. p 32

54. De Livera Alysha M, Hyndman Rob J, Snyder Ralph D (2011) Forecasting time series with complex seasonal patterns using exponential smoothing. J Am Stat Assoc 106(496):1513–1527

55. Deb Chirag, Zhang Fan, Yang Junjing, Lee Siew Eang, Shah Kwok Wei (2017) A review on time series forecasting techniques for building energy consumption. Renew Sustain Energy Rev 74:902–924

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

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

58. Díaz González F (2019) Federated learning for time series forecasting using lstm networks: exploiting similarities through clustering. Master's thesis, KTH Royal Institute of Technology, School of Electrical Engineering and Computer Science. http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-254665

59. Dong Jiaxiang, Wu Haixu, Zhang Haoran, Zhang Li, Wang Jianmin, Long Mingsheng (2023) Simmtm: A simple pre-training framework for masked time-series modeling. *arXiv preprint* [SPACE]arXiv:2302.00861. Accessed 2 Feb 2025

60. Dong Jiaxiang, Wu Haixu, Wang Yuxuan, Qiu Yunzhong, Zhang Li, Wang Jianmin, Long Mingsheng (2024) Timesiam: A pre-training framework for siamese time-series modeling. *arXiv preprint* [SPACE]arXiv:2402.02475. Accessed 2 Feb 2025

61. Dosovitskiy Alexey, Fischer Philipp, Springenberg Jost Tobias, Riedmiller Martin, Brox Thomas (2025) Discriminative unsupervised feature learning with exemplar convolutional neural networks. *arXiv preprint* [SPACE]arXiv:1406.6909. Accessed 2 Feb 2025

62. Drouin Alexandre, Marcotte Étienne, Chapados Nicolas (2022) Tactis: Transformer-attentional copulas for time series. In: International Conference on Machine Learning

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

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

65. Ekambaram Vijay, Jati Arindam, Nguyen Nam, Sinthong Phanwadee, Kalagnanam Jayant (2023) Tsmixer: Lightweight mlp-mixer model for multivariate time series forecasting. *arXiv preprint*[SPACE]arXiv:2306.09364. Accessed 2 Feb 2025

66. Eldele Emadeldeen, Ragab Mohamed, Chen Zhenghua, Wu Min, Kwoh Chee Keong, Li Xiaoli, Guan Cuntai (2021) Time-series representation learning via temporal and contextual contrasting. *arXiv preprint*[SPACE]arXiv:2106.14112. Accessed 2 Feb 2025

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

68. Eldele Emadeldeen, Ragab Mohamed, Chen Zhenghua, Wu Min, Li Xiaoli (2024) Tslanet: Rethinking transformers for time series representation learning. *arXiv preprint*[SPACE]arXiv:2404.08472. Accessed 2 Feb 2025

69. Enguehard Joseph (2023) Learning perturbations to explain time series predictions. *arXiv preprint*[SPACE]arXiv:2305.18840. Accessed 2 Feb 2025

70. Esteban Cristóbal, Hyland Stephanie L, Rätsch Gunnar (2017) Real-valued (medical) time series generation with recurrent conditional gans. *arXiv preprint*[SPACE]arXiv:1706.02633. Accessed 2 Feb 2025

71. Faloutsos Christos, Gasthaus Jan, Januschowski Tim, Wang Yuyang (2018) Forecasting big time series: old and new. Proc VLDB Endow 11(12):2102–2105

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

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

74. Fan Jianqing, Han Fang, Liu Han (2014) Challenges of big data analysis. Natl Sci Rev 1(2):293–314

75. Feng Cong, Chartan Erol Kevin, Hodge Bri-Mathias S, Zhang Jie (2017) Characterizing time series data diversity for wind forecasting. BDCAT. pp 113–119

76. Aya Ferchichi, Ben Abbes Ali, Vincent Barra, Manel Rhif, Riadh Farah Imed (2024) Multi-attention generative adversarial network for multi-step vegetation indices forecasting using multivariate time series. Eng Appl Artif Intell 128:107563

77. Folgado Duarte, Barandas Marília, Matias Ricardo, Martins Rodrigo, Carvalho Miguel, Gamboa Hugo (2018) Time alignment measurement for time series. Pattern Recogn 81:268–279

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

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

80. Fujimoto Scott, Hoof Herke, Meger David (2018) Addressing function approximation error in actor-critic methods. In: International Conference on Machine Learning. pp 1587–1596

81. Fuller Wayne A (2009) Introduction to statistical time series. John Wiley & Sons, New York

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

83. Gao, Jiechao, Wang, Wenpeng, Liu Zetian, Fazlay Rabbi Masum Billah Md, Campbell Bradford (2021) 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. pp 453–459

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

85. Gao Penglei, Yang Xi, Huang Kaizhu, Zhang Rui, Guo Ping, Goulermas John Y (2022) Egpde-net: Building continuous neural networks for time series prediction with exogenous variables. *arXiv preprint*[SPACE]arXiv:2208.01913. Accessed 2 Feb 2025

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

87. Gardner Jr Everette S (1985) Exponential smoothing: The state of the art. J Forecast 4(1):1–28

88. Garza Azul, Mergenthaler-Canseco Max (2023) Timegpt-1. *arXiv preprint*[SPACE]arXiv:2310.03589. Accessed 2 Feb 2025

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

90. Gilani, Faheem H (2021) *Diffusion maps and its applications to time series forecasting and filtering and second order elliptic PDEs*. The Pennsylvania State University

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

92. Godahewa Rakshitha, Bergmeir Christoph, Webb Geoff (2021) Pablo Montero-Manso. Rob Hyndman, Dominick dataset

93. Gong Zeying, Tang Yujin, Liang Junwei (2023) Patchmixer: A patch-mixing architecture for long-term time series forecasting. *arXiv preprint*[SPACE]arXiv:2310.00655. Accessed 2 Feb 2025

94. Goodfellow Ian, Bengio Yoshua (2016) Aaron Courville. MIT Press, Deep learning

95. Gruver Nate, Finzi Marc, Qiu Shikai, Wilson Andrew Gordon (2023) Large language models are zero-shot time series forecasters. *arXiv preprint*[SPACE]arXiv:2310.07820. Accessed 2 Feb 2025

96. Gu Albert, Goel Karan, Ré Christopher (2021) Efficiently modeling long sequences with structured state spaces. *arXiv preprint*[SPACE]arXiv:2111.00396. Accessed 2 Feb 2025

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

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

99. Guo Wenzhong, Wang Jianwen, Wang Shiping (2019) Deep multimodal representation learning: a survey. Ieee Access 7:63373–63394

100. Haarnoja Tuomas, Zhou Aurick, Hartikainen Kristian, Tucker George, Ha Sehoon, Tan Jie, Kumar Vikash, Zhu Henry, Gupta

Abhishek, Abbeel Pieter, et al (2018) Soft actor-critic algorithms and applications. *arXiv preprint*[SPACE]arXiv:1812.05905. Accessed 2 Feb 2025

101. Hamilton James D (2020) Time series analysis. Princeton University Press, Princeton

102. Hamzaçebi Coşkun (2008) Improving artificial neural networks' performance in seasonal time series forecasting. Inf Sci 178(23):4550–4559

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

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

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

106. He Xiaoyu, Shi Suixiang, Geng Xiulin, Lingyu Xu (2022) Dynamic co-attention networks for multi-horizon forecasting in multivariate time series. Futur Gener Comput Syst 135:72–84

107. He Xiaoyu, Shi Suixiang, Geng Xiulin, Jie Yu, Lingyu Xu (2023) Multi-step forecasting of multivariate time series using multi-attention collaborative network. Expert Syst Appl 211:118516

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

109. Hipel Keith W, Ian McLeod A (1994) Time series modelling of water resources and environmental systems. Elsevier, Amsterdam

110. Ma Hiransha, Ab Gopalakrishnan E, Krishna Menon Vijay, Soman KP (2018) Nse stock market prediction using deep-learning models. Proc Comput Sci 132:1351–1362

111. Ho Tin Kam (1995) Random decision forests. Proc 3rd Int Conf Document Anal Recogn. 1:278–282

112. Hu Jiaxi, Lan Disen, Zhou Ziyu, Wen Qingsong, Liang Yuxuan (2024) Time-ssm: simplifying and unifying state space models for time series forecasting. *arXiv preprint*[SPACE]arXiv:2405.16312. Accessed 2 Feb 2025

113. Hyndman RJ, Athanasopoulos G (2018) *Forecasting: Principles and Practice*. OTexts

114. Ilbert Romain, Odonnat Ambroise, Feofanov Vasilii, Virmaux Aladin, Paolo Giuseppe, Palpanas Themis, Redko Ievgen (2024) Unlocking the potential of transformers in time series forecasting with sharpness-aware minimization and channel-wise attention. *arXiv preprint*[SPACE]arXiv:2402.10198. Accessed 2 Feb 2025

115. Januschowski Tim, Gasthaus Jan, Wang Yuyang, Salinas David, Flunkert Valentin, Bohlke-Schneider Michael, Callot Laurent (2020) Criteria for classifying forecasting methods. Int J Forecast 36(1):167–177

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

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

118. Jin Ming, Wang Shiyu, Ma Lintao, Chu Zhixuan, Zhang James Y, Shi Xiaoming, Chen Pin-Yu, Liang Yuxuan, Li Yuan-Fang, Pan Shirui, et al (2023) Time-llm: Time series forecasting by reprogramming large language models. *arXiv preprint*[SPACE]arXiv:2310.01728. Accessed 2 Feb 2025

119. Kalra Riya, Singh Tinku, Mishra Suryanshi, Kumar Naveen, Kim Taehong, Kumar Manish et al (2024) An efficient hybrid approach for forecasting real-time stock market indices. J King Saud Univ-Comput Inf Sci 36(8):102180

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

121. Klopries Hendrik, Schwung Andreas (2024) Itf-gan: synthetic time series dataset generation and manipulation by interpretable features. Knowl-Based Syst 283:111131

122. Koc Erdinc, Türkoğlu Muammer (2022) 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, pp 1–9

123. Koelstra Sander, Muhl Christian, Soleymani Mohammad, Lee Jong-Seok, Yazdani Ashkan, Ebrahimi Touradj, Pun Thierry, Nijholt Anton, Patras Ioannis (2011) Deap: a database for emotion analysis; using physiological signals. IEEE Trans Affect Comput 3(1):18–31

124. Koh Pang Wei, Liang Percy (2017) Understanding black-box predictions via influence functions. In: International Conference on Machine Learning. PMLR pp 1885–1894

125. Kolassa Stephan (2020) Why the "best'' point forecast depends on the error or accuracy measure. Int J Forecast 36(1):208–211

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

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

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

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

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

131. Nikolaos Kourentzes, Fotios Petropoulos, Trapero Juan R (2014) Improving forecasting by estimating time series structural components across multiple frequencies. Int J Forecast 30(2):291–302

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

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

134. LeCun Yann, Bengio Yoshua, Hinton Geoffrey (2015) Deep learning. Nature 521(7553):436–444

135. Lee Junsoo (1994) Univariate time series modeling and forecasting (box-jenkins method). Econ Times. p 413

136. Sang Il Lee and Seong Joon Yoo (2020) Threshold-based portfolio: the role of the threshold and its applications. J Supercomput 76(10):8040–8057

137. Li Jun, Liu Che, Cheng Sibo, Arcucci Rossella, Hong Shenda (2023) Frozen language model helps ecg zero-shot learning. *arXiv preprint*[SPACE]arXiv:2303.12311. Accessed 2 Feb 2025

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

139. Li Rui, Shahn Zach, Li Jun, Lu Mingyu, Chakraborty Prithwish, Sow Daby, Ghalwash Mohamed, Lehman Li-wei H (2020) G-net: a deep learning approach to g-computation for counterfactual outcome prediction under dynamic treatment regimes. *arXiv preprint*[SPACE]arXiv:2003.10551. Accessed 2 Feb 2025

140. Li Shancang, Da Li Xu, Zhao Shanshan (2015) The internet of things: a survey. Inf Syst Front 17:243–259

141. Li S, Jin X, Xuan Y, Zhou X, Chen W, Wang YX, Yan X (2019) Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. Adv Neural Inf Process Syst 32:11

142. Li Tong, Liu Zhaoyang, Shen Yanyan, Wang Xue, Chen Haokun, Huang Sen (2024) Master: market-guided stock transformer for stock price forecasting. Proc AAAI Conf Artif Intell 38:162–170

143. Li Xuerong, Shang Wei, Wang Shouyang (2019) Text-based crude oil price forecasting: a deep learning approach. Int J Forecast 35(4):1548–1560

144. Li Yaguang, Yu Rose, Shahabi Cyrus, Liu Yan (2017) Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv preprint*[SPACE]arXiv:1707.01926. Accessed 2 Feb 2025

145. Li Yan, Xinjiang Lu, Wang Yaqing, Dou Dejing (2022) Generative time series forecasting with diffusion, denoise, disentanglement. Adv Neural Inf Process Syst 35:23009–23022

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

147. Li Zhe, Rao Zhongwen, Pan Lujia, Wang Pengyun, Xu Zenglin (2023) Ti-mae: Self-supervised masked time series autoencoders. *arXiv preprint*[SPACE]arXiv:2301.08871. Accessed 2 Feb 2025

148. Lillicrap Timothy P, Hunt Jonathan J, Pritzel Alexander, Heess Nicolas, Erez Tom, Tassa Yuval, Silver David, Wierstra Daan (2015) Continuous control with deep reinforcement learning. *arXiv preprint*[SPACE]arXiv:1509.02971. Accessed 2 Feb 2025

149. Lim Bryan (2018) Forecasting treatment responses over time using recurrent marginal structural networks. Adv Neural Inf Process Syst. p 31

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

151. Lin Shengsheng, Lin Weiwei, Wu Wentai, Wang Songbo, Wang Yongxiang (2023) Petformer: Long-term time series forecasting via placeholder-enhanced transformer. *arXiv preprint*[SPACE]arXiv:2308.04791. Accessed 2 Feb 2025

152. Lin Wen-Hui, Wang Ping, Chao Kuo-Ming, Lin Hsiao-Chung, Yang Zong-Yu, Lai Yu-Huang (2021) Wind power forecasting with deep learning networks: time-series forecasting. Appl Sci 11(21):10335

153. Linot Alec J, Burby Joshua W, Tang Qi, Balaprakash Prasanna, Graham Michael D, Maulik Romit (2023) Stabilized neural ordinary differential equations for long-time forecasting of dynamical systems. J Comput Phys 474:111838

154. Liu Mingzhou, Sun Xinwei, Hu Lingjing, Wang Yizhou (2023) Causal discovery from subsampled time series with proxy variables. *arXiv preprint*[SPACE]arXiv:2305.05276. Accessed 2 Feb 2025

155. Liu Minhao, Zeng Ailing, Chen Muxi, Zhijian Xu, Lai Qiuxia, Ma Lingna, Qiang Xu (2022) Scinet: time series modeling and forecasting with sample convolution and interaction. Adv Neural Inf Process Syst 35:5816–5828

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

157. Liu Xin, McDuff Daniel, Kovacs Geza, Galatzer-Levy Isaac, Sunshine Jacob, Zhan Jiening, Poh Ming-Zher, Liao Shun, Di Achille Paolo, Patel Shwetak (2023) Large language models are few-shot health learners. *arXiv preprint*[SPACE]arXiv:2305.15525. Accessed 2 Feb 2025

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

159. Liu Yong, Haixu Wu, Wang Jianmin, Long Mingsheng (2022) Non-stationary transformers: exploring the stationarity in time series forecasting. Adv Neural Inf Process Syst 35:9881–9893

160. Liu Yong, Hu Tengge, Zhang Haoran, Wu Haixu, Wang Shiyu, Ma Lintao, Long Mingsheng (2023) itransformer: Inverted transformers are effective for time series forecasting. *arXiv preprint*[SPACE]arXiv:2310.06625. Accessed 2 Feb 2025

161. Lundberg Scott M, Lee Su-In (2017) A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. p 30

162. Luo Dongsheng, Cheng Wei, Wang Yingheng, Dongkuan Xu, Ni Jingchao, Wenchao Yu, Zhang Xuchao, Liu Yanchi, Chen Yuncong, Chen Haifeng et al (2023) Time series contrastive learning with information-aware augmentations. Proc AAAI Conf Artif Intell 37:4534–4542

163. Luo Rui, Zhang Weinan, Xu Xiaojun, Wang Jun (2018) A neural stochastic volatility model. In: Proceedings of the AAAI Conference on Artificial Intelligence. p 32

164. Lütkepohl Helmut (2005) Vector autoregressive moving average processes. New Introduction to Multiple Time Series Analysis. pp 419–446

165. Lv Yisheng, Duan Yanjie, Kang Wenwen, Li Zhengxi, Wang Fei-Yue (2014) Traffic flow prediction with big data: a deep learning approach. IEEE Trans Intell Transp Syst 16(2):865–873

166. Lyu Xinrui, Hueser Matthias, Hyland Stephanie L, Zerveas George, Raetsch Gunnar (2018) Improving clinical predictions through unsupervised time series representation learning. *arXiv preprint*[SPACE]arXiv:1812.00490. Accessed 2 Feb 2025

167. Makridakis Spyros (1978) Time-series analysis and forecasting: an update and evaluation. International Statistical Review/Revue Internationale de Statistique. pp 255–278

168. Makridakis Spyros, Spiliotis Evangelos, Assimakopoulos Vassilios (2018) The m4 competition: results, findings, conclusion and way forward. Int J Forecast 34(4):802–808

169. Makridakis Spyros, Spiliotis Evangelos, Assimakopoulos Vassilios (2018) Statistical and machine learning forecasting methods: concerns and ways forward. PLoS ONE 13(3):e0194889

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

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

172. Mehrkanoon Siamak (2019) Deep shared representation learning for weather elements forecasting. Knowl-Based Syst 179:120–128

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

174. Montero-Manso Pablo, Hyndman Rob J (2021) Principles and algorithms for forecasting groups of time series: locality and globality. Int J Forecast 37(4):1632–1653

175. Montgomery Douglas C, Jennings Cheryl L, Murat Kulahci (2015) Introduction to time series analysis and forecasting. John Wiley & Sons, New York

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

177. Mudelsee Manfred (2019) Trend analysis of climate time series: a review of methods. Earth Sci Rev 190:310–322

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

179. Ni Zelin, Yu Hang, Liu Shizhan, Li Jianguo, Lin Weiyao (2023) Basisformer: Attention-based time series forecasting with learnable and interpretable basis. *arXiv preprint* [SPACE]arXiv:2310.20496. Accessed 2 Feb 2025

180. Nie Yuqi, Nguyen Nam H, Sinthong Phanwadee, Kalagnanam Jayant (2022) A time series is worth 64 words: Long-term forecasting with transformers. *arXiv preprint* [SPACE]arXiv:2211.14730. Accessed 2 Feb 2025

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

182. O'Hara Maureen, Ye Mao (2011) Is market fragmentation harming market quality? J Financ Econ 100(3):459–474

183. Oord Aaron van den, Dieleman Sander, Zen Heiga, Simonyan Karen, Vinyals Oriol, Graves Alex, Kalchbrenner Nal, Senior Andrew, Kavukcuoglu Koray (2016) Wavenet: A generative model for raw audio. *arXiv preprint* [SPACE]arXiv:1609.03499. Accessed 2 Feb 2025

184. van den Oord Aaron, Li Yazhe, Vinyals Oriol (2018) Representation learning with contrastive predictive coding. *arXiv preprint* [SPACE]arXiv:1807.03748. Accessed 2 Feb 2025

185. Oreshkin Boris N, Carpov Dmitri, Chapados Nicolas, Bengio Yoshua (2019) N-beats: Neural basis expansion analysis for interpretable time series forecasting. *arXiv preprint* [SPACE]arXiv:1905.10437. Accessed 2 Feb 2025

186. Ahmed Oussous, Fatima-Zahra Benjelloun, Ait Lahcen Ayoub, Samir Belfkih (2018) Big data technologies: a survey. J King Saud Univ-Comput Inf Sci 30(4):431–448

187. Ozyurt Yilmazcan, Feuerriegel Stefan, Zhang Ce (2022) Contrastive learning for unsupervised domain adaptation of time series. *arXiv preprint* [SPACE]arXiv:2206.06243. Accessed 2 Feb 2025

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

189. Peng Hao, Yang Renyu, Wang Zheng, Li Jianxin, Lifang He SYu, Philip Albert Y, Zomaya Rajiv Ranjan (2021) Lime: low-cost and incremental learning for dynamic heterogeneous information networks. IEEE Trans Comput 71(3):628–642

190. Perslev Mathias, Jensen Michael, Darkner Sune, Jennum Poul Jørgen, Igel Christian (2019) U-time: A fully convolutional network for time series segmentation applied to sleep staging. Adv Neural Inf Process Syst. p 32

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

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

193. Qin Yao, Song Dongjin, Chen Haifeng, Cheng Wei, Jiang Guofei, Cottrell Garrison (2017) A dual-stage attention-based recurrent neural network for time series prediction. *arXiv preprint* [SPACE]arXiv:1704.02971. Accessed 2 Feb 2025

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

195. Rasp Stephan, Dueben Peter D, Scher Sebastian, Weyn Jonathan A, Mouatadid Soukayna, Thuerey Nils (2020) Weatherbench: a benchmark data set for data-driven weather forecasting. J Adv Model Earth Syst 12(11):e2020MS002203

196. Rasul Kashif, Seward Calvin, Schuster Ingmar, Vollgraf Roland (2021) Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In: International Conference on Machine Learning. PMLR. pp 8857–8868.

197. Rasul Kashif, Ashok Arjun, Williams Andrew Robert, Khorasani Arian, Adamopoulos George, Bhagwatkar Rishika, Biloš Marin, Ghonia Hena, Hassen Nadhir Vincent, Schneider Anderson, et al. (2023) Lag-llama: Towards foundation models for time series forecasting. *arXiv preprint* [SPACE]arXiv:2310.08278. Accessed 2 Feb 2025

198. Rebjock Quentin, Kurt Baris, Januschowski Tim, Callot Laurent (2021) Online false discovery rate control for anomaly detection in time series. Adv Neural Inf Process Syst 34:26487–26498

199. Ribeiro Marco Tulio, Singh Sameer, Guestrin Carlos (2016) " 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. pp 1135–1144

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

201. Rokach Lior (2016) Decision forest: twenty years of research. Inf Fusion 27:111–125

202. Olaf Ronneberger, Philipp Fischer, Thomas Brox (2015) U-net: convolutional networks for biomedical image segmentation. Medical image computing and computer-assisted intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18. Springer, New York, pp 234–241

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

204. Ruder S (2017) An overview of multi-task learning in deep neural networks. *arXiv preprint* [SPACE]arXiv:1706.05098. Accessed 2 Feb 2025

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

206. Salinas David, Flunkert Valentin, Gasthaus Jan, Januschowski Tim (2020) Deepar: probabilistic forecasting with autoregressive recurrent networks. Int J Forecast 36(3):1181–1191

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

208. Savi Marco, Olivadese Fabrizio (2021) Short-term energy consumption forecasting at the edge: a federated learning approach. IEEE Access 9:95949–95969

209. Saxena Harshit, Aponte Omar, McConky Katie T (2019) A hybrid machine learning model for forecasting a billing period's peak electric load days. Int J Forecast 35(4):1288–1303

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

211. Semenoglou Artemios-Anargyros, Spiliotis Evangelos, Makridakis Spyros, Assimakopoulos Vassilios (2021) Investigating the accuracy of cross-learning time series forecasting methods. Int J Forecast 37(3):1072–1084

212. Seyfi Ali, Rajotte Jean-Francois, Ng Raymond (2022) Generating multivariate time series with common source coordinated gan (cosci-gan). Adv Neural Inf Process Syst 35:32777–32788

213. Omer Berat Sezer (2020) Mehmet Ugur Gudelek, Ahmet Murat Ozbayoglu (2020) Financial time series forecasting with deep learning A systematic literature review: 2005–2019. Appl Soft Comput. 90:106181.

214. Shabani, Amin, Abdi, Amir, Meng, Lili, Sylvain, Tristan (2022) Scaleformer: iterative multi-scale refining transformers for time series forecasting. *arXiv preprint*[SPACE]arXiv:2206.04038. Accessed 2 Feb 2025

215. Shao Zezhi, Zhang Zhao, Wang Fei (2022) 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. pp 1567–1577

216. Shen Lifeng, Kwok James (2020) Non-autoregressive conditional diffusion models for time series prediction. *arXiv preprint*[SPACE]arXiv:2306.05043. Accessed 2 Feb 2025

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

218. Shi Xingjian, Chen Zhourong, Wang Hao, Yeung Dit-Yan, Wong Wai-Kin, Woo Wang-chun (2020) Convolutional lstm network A machine learning approach for precipitation nowcasting. Adv Neural Inf Process Syst. p 28

219. Shumway Robert H, Stoffer David S, Stoffer David S (2000) Time series analysis and its applications. Springer, Cham

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

221. Simonyan Karen, Vedaldi Andrea, Zisserman Andrew (2013) Deep inside convolutional networks: visualising image classification models and saliency maps. *arXiv preprint*[SPACE]arXiv:1312.6034. Accessed 2 Feb 2025

222. Singh Ankit (2021) Clda: contrastive learning for semi-supervised domain adaptation. Adv Neural Inf Process Syst 34:5089–5101

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

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

225. Smyl Slawek (2020) A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. Int J Forecast 36(1):75–85

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

227. Subramanya Tejas, Riggio Roberto (2021) Centralized and federated learning for predictive vnf autoscaling in multi-domain 5g networks and beyond. IEEE Trans Netw Serv Manage 18(1):63–78

228. Sun Chenxi, Li Yaliang, Li Hongyan, Hong Shenda (2023) Test: Text prototype aligned embedding to activate llm's ability for time series. *arXiv preprint*[SPACE]arXiv:2308.08241. Accessed 2 Feb 2025

229. Sun Fan-Keng, Boning Duane S (2022) Fredo: frequency domain-based long-term time series forecasting. *arXiv preprint*[SPACE]arXiv:2205.12301. Accessed 2 Feb 2025

230. Sun Fan-Keng, Lang Chris, Boning Duane (2021) Adjusting for autocorrelated errors in neural networks for time series. Adv Neural Inf Process Syst 34:29806–29819

231. Sutskever I (2014) Sequence to sequence learning with neural networks. *arXiv preprint*[SPACE]arXiv:1409.3215. Accessed 2 Feb 2025

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

233. Shuntaro Takahashi Yu, Chen Kumiko Tanaka-Ishii (2019) Modeling financial time-series with generative adversarial networks. Physica A 527:121261

234. Tang Peiwang, Zhang Xianchao (2023) Infomaxformer: Maximum entropy transformer for long time-series forecasting problem. *arXiv preprint*[SPACE]arXiv:2301.01772. Accessed 2 Feb 2025

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

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

237. Tedjopurnomo David Alexander, Bao Zhifeng, Zheng Baihua, Choudhury Farhana Murtaza, Qin Alex Kai (2020) A survey on modern deep neural network for traffic prediction: trends, methods and challenges. IEEE Trans Knowl Data Eng 34(4):1544–1561

238. Topol Eric J (2019) High-performance medicine: the convergence of human and artificial intelligence. Nat Med 25(1):44–56

239. Torres José F, Troncoso Alicia, Koprinska Irena, Wang Zheng, Martínez-Álvarez Francisco (2019) 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*. Springer, Cham. pp 123–133

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

241. Trindade Artur (2015) Electricityloaddiagrams 2011-2014. UCI Machine Learning Repository

242. Wang Hao, Zhang Zhenguo (2022) 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). IEEE. pp 26–31.

243. Wang Lei, Zeng Liang, Li Jian (2023) Aec-gan: adversarial error correction gans for auto-regressive long time-series generation. Proc AAAI Conf Artifi Intell 37:10140–10148

244. Wang Shiyu, Wu Haixu, Shi Xiaoming, Hu Tengge, Luo Huakun, Ma Lintao, Zhang James Y, ZHOU JUN (2024) Timemixer:

Decomposable multiscale mixing for time series forecasting. In: International Conference on Learning Representations (ICLR).

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

246. Wang Yuxuan, Wu Haixu, Dong Jiaxiang, Liu Yong, Qiu Yunzhong, Zhang Haoran, Wang Jianmin, Long Mingsheng (2024) Timexer: Empowering transformers for time series forecasting with exogenous variables. *arXiv preprint*[SPACE]arXiv:2402.19072. Accessed 2 Feb 2025

247. Wang Zhiyuan, Xovee Xu, Zhang Weifeng, Trajcevski Goce, Zhong Ting, Zhou Fan (2022) Learning latent seasonal-trend representations for time series forecasting. Adv Neural Inf Process Syst 35:38775–38787

248. Wen Ruofeng, Torkkola Kari, Narayanaswamy Balakrishnan, Madeka Dhruv (2017) A multi-horizon quantile recurrent forecaster. *arXiv preprint*[SPACE]arXiv:1711.11053. Accessed 2 Feb 2025

249. Woo Gerald, Liu Chenghao, Sahoo Doyen, Kumar Akshat, Hoi Steven (2022) Cost: Contrastive learning of disentangled seasonal-trend representations for time series forecasting. *arXiv preprint*[SPACE]arXiv:2202.01575. Accessed 2 Feb 2025

250. Woo Gerald, Liu Chenghao, Sahoo Doyen, Kumar Akshat, Hoi Steven (2022) Etsformer: Exponential smoothing transformers for time-series forecasting. *arXiv preprint*[SPACE]arXiv:2202.01381. Accessed 2 Feb 2025

251. Haixu Wu, Jiehui Xu, Wang Jianmin, Long Mingsheng (2021) Autoformer: decomposition transformers with auto-correlation for long-term series forecasting. Adv Neural Inf Process Syst 34:22419–22430

252. Wu Haixu, Hu Tengge, Liu Yong, Zhou Hang, Wang Jianmin, Long Mingsheng (2022) Timesnet: temporal 2d-variation modeling for general time series analysis. *arXiv preprint*[SPACE]arXiv:2210.02186. Accessed 2 Feb 2025

253. Wu Neo, Green Bradley, Ben Xue, O'Banion Shawn (2020) Deep transformer models for time series forecasting: The influenza prevalence case. *arXiv preprint*[SPACE]arXiv:2001.08317. Accessed 2 Feb 2025

254. Sifan Wu, Xiao Xi, Ding Qianggang, Zhao Peilin, Wei Ying, Huang Junzhou (2020) Adversarial sparse transformer for time series forecasting. Adv Neural Inf Process Syst 33:17105–17115

255. Xindong Wu, Zhu Xingquan, Gong-Qing Wu, Ding Wei (2013) Data mining with big data. IEEE Trans Knowl Data Eng 26(1):97–107

256. Xie Qianqian, Han Weiguang, Lai Yanzhao, Peng Min, Huang Jimin (2023) The wall street neophyte: A zero-shot analysis of chatgpt over multimodal stock movement prediction challenges. *arXiv preprint*[SPACE]arXiv:2304.05351. Accessed 2 Feb 2025

257. Qifa Xu, Liu Xi, Jiang Cuixia, Keming Yu (2016) Quantile autoregression neural network model with applications to evaluating value at risk. Appl Soft Comput 49:1–12

258. Yingcheng Xu, Zhang Yunfeng, Liu Peide, Zhang Qiuyue, Zuo Yuqi (2024) Gan-enhanced nonlinear fusion model for stock price prediction. Int J Comput Intell Syst 17(1):12

259. Xue H, Salim FD (2023) Promptcast: a new prompt-based learning paradigm for time series forecasting. IEEE Trans Knowl Data Eng 36:1–14

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

261. Xue Wang, Zhou Tian, Wen QingSong, Gao Jinyang, Ding Bolin, Jin Rong (2023) Make transformer great again for time series forecasting: Channel aligned robust dual transformer. *arXiv preprint*[SPACE]arXiv:2305.12095. Accessed 2 Feb 2025

262. Vijaya Krishna Yalavarthi, Kiran Madhusudhanan, Randolf Scholz, Nourhan Ahmed, Johannes Burchert, Shayan Jawed, Stefan Born, Lars Schmidt-Thieme. Grafiti (2024) Graphs for forecasting irregularly sampled time series. Proc AAAI Conf Artif Intell. 38:16255–16263

263. Yan Jingquan, Wang Hao (2023) Self-interpretable time series prediction with counterfactual explanations. *arXiv preprint*[SPACE]arXiv:2306.06024. Accessed 2 Feb 2025

264. Hao-Fan Yang and Yi-Ping Phoebe Chen (2019) Representation learning with extreme learning machines and empirical mode decomposition for wind speed forecasting methods. Artif Intell 277:103176

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

266. Yi Kun, Zhang Qi, Fan Wei, Wang Shoujin, Wang Pengyang, He Hui, Lian Defu, An Ning, Cao Longbing, Niu Zhendong (2023) Frequency-domain mlps are more effective learners in time series forecasting. *arXiv preprint*[SPACE]arXiv:2311.06184. Accessed 2 Feb 2025

267. Ozge Cagcag Yolcu and Ufuk Yolcu (2023) A novel intuitionistic fuzzy time series prediction model with cascaded structure for financial time series. Expert Syst Appl 215:119336

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

269. Jinsung Yoon, Daniel Jarrett, Mihaela Van der Schaar (2019) Time-series generative adversarial networks. Adv Neural Inf Process Syst. 32

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

271. Yu Bing, Yin Haoteng, Zhu Zhanxing (2017) Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv preprint*[SPACE]arXiv:1709.04875. Accessed 2 Feb 2025

272. Yu Chengqing, Wang Fei, Shao Zezhi, Qian Tangwen, Zhang Zhao, Wei Wei, Xu Yongjun (2024) 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. pp 3989–4000

273. Yu Xinli, Chen Zheng, Ling Yuan, Dong Shujing, Liu Zongyi, Lu Yanbin (2023) Temporal data meets llm–explainable financial time series forecasting. *arXiv preprint*[SPACE]arXiv:2306.11025. Accessed 2 Feb 2025

274. Yue Zhihan, Wang Yujing, Duan Juanyong, Yang Tianmeng, Huang Congrui, Tong Yunhai, Bixiong Xu (2022) Ts2vec: towards universal representation of time series. Proc AAAI Conf Artif Intell 36:8980–8987

275. Zaini Nur'atiah, Ean Lee Woen, Ahmed Ali Najah, Malek Marlinda Abdul (2022) A systematic literature review of deep learning neural network for time series air quality forecasting. Environ Sci Pollut Res. pp 1–33

276. Zebari Rizgar, Abdulazeez Adnan, Zeebaree Diyar, Zebari Dilovan, Saeed Jwan (2020) A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. J Appl Sci Technol Trends 1(1):56–70

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

278. Zhang Boyu, Yang Hongyang, Liu Xiao-Yang (2023) Instruct-fingpt: Financial sentiment analysis by

instruction tuning of general-purpose large language models. *arXiv preprint*[SPACE]arXiv:2306.12659. Accessed 2 Feb 2025

279. Zhang Chenhan, Shuyu Zhang JQ, James Shui Yu (2021) Fast-gnn: a topological information protected federated learning approach for traffic speed forecasting. IEEE Trans Industr Inf 17(12):8464–8474

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

281. Guoqiang Zhang B, Patuwo Eddy, Hu Michael Y (1998) Forecasting with artificial neural networks: the state of the art. Int J Forecast 14(1):35–62

282. Zhang Wenrui, Yang Ling, Geng Shijia, Hong Shenda (2022) Self-supervised time series representation learning via cross reconstruction transformer. *arXiv preprint*[SPACE]arXiv:2205.09928. Accessed 2 Feb 2025

283. Zhang Xiaoning, Fang Fang, Wang Jiaqi (2020) Probabilistic solar irradiation forecasting based on variational bayesian inference with secure federated learning. IEEE Trans Industr Inf 17(11):7849–7859

284. Zhang Xiyuan, Jin Xiaoyong, Gopalswamy Karthick, Gupta Gaurav, Park Youngsuk, Shi Xingjian, Wang Hao, Maddix Danielle C, Wang Yuyang (2022) First de-trend then attend: Rethinking attention for time-series forecasting. *arXiv preprint*[SPACE]arXiv:2212.08151. Accessed 2 Feb 2025

285. Zhang Yifan, Wu Rui, Dascalu Sergiu M, Harris Jr Frederick C (2023) Multi-scale transformer pyramid networks for multivariate time series forecasting. *arXiv preprint*[SPACE]arXiv:2308.11946. Accessed 2 Feb 2025

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

287. Zhang Zhenwei, Wang Xin, Gu Yuantao (2023) Sageformer: Series-aware graph-enhanced transformers for multivariate time series forecasting. *arXiv preprint*[SPACE]arXiv:2307.01616. Accessed 2 Feb 2025

288. Zhang Zhenwei, Meng Linghang, Yuantao Gu (2024) Sageformer: series-aware framework for long-term multivariate time series forecasting. IEEE Internet Things J. https://doi.org/10.1109/JIOT.2024.3363451

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

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

291. Zheng Xiaochen, Chen Xingyu, Schürch Manuel, Mollaysa Amina, Allam Ahmed, Krauthammer Michael (2023) Simts: Rethinking contrastive representation learning for time series forecasting. *arXiv preprint*[SPACE]arXiv:2303.18205. Accessed 2 Feb 2025

292. Zhou Haoyi, Zhang Shanghang, Peng Jieqi, Zhang Shuai, Li Jianxin, Xiong Hui, Zhang Wancai (2021) Informer: beyond efficient transformer for long sequence time-series forecasting. Proc AAAI Conf Artif Intell 35:11106–11115

293. Zhou Tian, Ma Ziqing, Wen Qingsong, Sun Liang, Yao Tao, Yin Wotao, Jin Rong et al (2022) Film: frequency improved legendre memory model for long-term time series forecasting. Adv Neural Inf Process Syst 35:12677–12690

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

295. Zhou Tian, Zhu Jianqing, Wang Xue, Ma Ziqing, Wen Qingsong, Sun Liang, Jin Rong (2022) Treedrnet: a robust deep model for long term time series forecasting. *arXiv preprint*[SPACE]arXiv:2206.12106. Accessed 2 Feb 2025

296. Zhou Tian, Niu Peisong, Wang Xue, Sun Liang, Jin Rong (2023) One fits all: Power general time series analysis by pretrained lm. *arXiv preprint*[SPACE]arXiv:2302.11939. Accessed 2 Feb 2025

297. Zhu Hongjun, Yuan Shun, Liu Xin, Chen Kuo, Jia Chaolong, Qian Ying (2024) Casciff: A cross-domain information fusion framework tailored for cascade prediction in social networks. Knowl-Based Syst. p 112391

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

299. Zou Dongcheng, Wang Senzhang, Li Xuefeng, Peng Hao, Wang Yuandong, Liu Chunyang, Sheng Kehua, Zhang Bo (2024) 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. pp 1032–1041