Technology investigation on time series classification and prediction

Yuerong Tong[1], Jingyi Liu[1], Lina Yu[1], Liping Zhang[1], Linjun Sun[1], Weijun Li[1,2], Xin Ning[1], Jian Xu[1], Hong Qin[1] and Qiang Cai[3]

[1] Institute of Semiconductors, Chinese Academy of Sciences, Beijing, China
[2] Shenzhen DAPU Microelectronics Co., Ltd., Shenzhen, China
[3] National Engineering Laboratory for Agri-product Quality Traceability, Beijing Technology and Business University, Beijing, China

ABSTRACT

Time series appear in many scientific fields and are an important type of data. The use of time series analysis techniques is an essential means of discovering the knowledge hidden in this type of data. In recent years, many scholars have achieved fruitful results in the study of time series. A statistical analysis of 120,000 literatures published between 2017 and 2021 reveals that the topical research about time series is mostly focused on their classification and prediction. Therefore, in this study, we focus on analyzing the technical development routes of time series classification and prediction algorithms. 87 literatures with high relevance and high citation are selected for analysis, aiming to provide a more comprehensive reference base for interested researchers. For time series classification, it is divided into supervised methods, semi-supervised methods, and early classification of time series, which are key extensions of time series classification tasks. For time series prediction, from classical statistical methods, to neural network methods, and then to fuzzy modeling and transfer learning methods, the performance and applications of these different methods are discussed. We hope this article can help aid the understanding of the current development status and discover possible future research directions, such as exploring interpretability of time series analysis and online learning modeling.

Corresponding authors
Lina Yu, yulina@semi.ac.cn
Qiang Cai, caiq@th.btbu.edu.cn

INTRODUCTION

Time series are a set of observations made and recorded at different points in time (*Misra & Siddharth, 2017*). It is ubiquitous in real life. Whether measured during natural processes (weather, sound waves) or artificially generated processes (stock, robots), most real-world data contain time elements (*Langkvist, Karlsson & Loutfi, 2014*). Moreover, time series data are being produced in different fields at an unprecedented scale and speed. Therefore, knowledge discovery from time series has considerable potential. Because of its unique sequence characteristic, time series analysis is considered one of the ten most-challenging problems in the field of data mining (*Yang & Wu, 2006*), becoming a prevalent research topic that has attracted the attention of many researchers over the years (*Schreiber, 1999*; *Osmanoglu et al., 2016*).
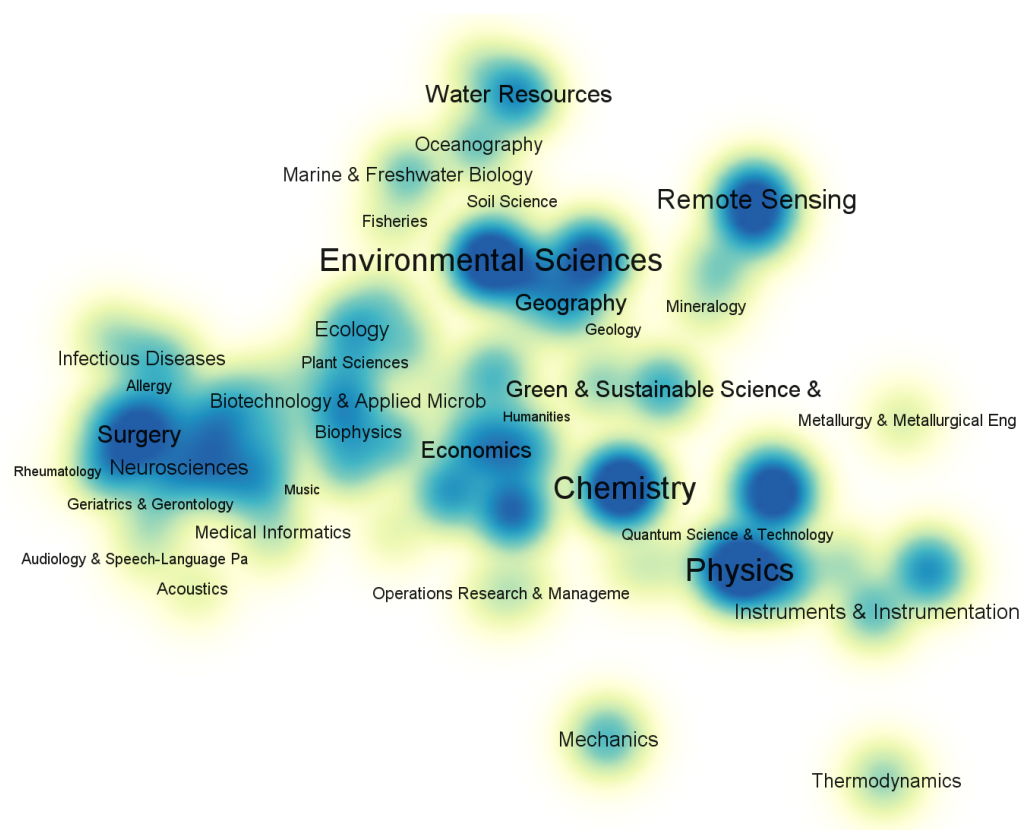
Figure 1 **Subject category co-occurrence map of time series literatures (2017–2021).**

In time series analysis, common data sets are often used, such as UCR time series classification archive (https://www.cs.ucr.edu/~eamonn/time_series_data_2018/), Awesome Public Dataset (https://github.com/awesomedata/awesome-public-datasets) and CEIC (https://www.ceicdata.com/zh-hans).

To gain a comprehensive understanding of the current status of time series application, we use time series as a keyword to search the Web of Science Core Collection and collect 120,000 references published between 2017 and 2021. Then, we use VOSViewer (Leiden University, The Netherlands) to visualize anaysis result: the subject category co-occurrence map of first-level disciplines as shown in Fig. 1.

To gain a clearer understanding of the application fields of time series, we remove the two subjects with the highest number of matches, *i.e.,* Engineering and Computer Science, both of which have a high total link strength; this can be attributed to the fact that these two subjects are often used in research as analysis tools for other domains. The 120,000 publications contain 161 unique level-1 subjects in total. From Fig. 1, we can see that time series has an extensive range of applications.

Time series has been widely used in many fields (*Jiang, 2015*) such as environmental sciences (*Shahane, Thomas & Bock, 1977*; *Gluhovsky & Agee, 2007*), chemistry (*Bates et al., 2014*), physics (*Song & Russell, 1999*), mathematics (*Corradi, 1995*; *McDonald & Alan,*

1986), biomedical (*Bar-Joseph, 2004*; *Zeger, Irizarry & Peng, 2006*), meteorology (*Ghil et al., 2002*), astronomy (*Zhu, 2017*), finance (*Sezer, Gudelek & Ozbayoglu, 2020*), and other fields (*Li et al., 2019*). Specifically, in medicine, medical time series test data can be used to diagnose diseases (for example, diagnoses of heart disease *Kadous & Sammut, 2005*), predict disease counts (*Talaei-Khoei & Wilson, 2019*), evaluate the impact of interventions on public health over time (*Lopez, Steven & Antonio, 2017*), and analyze gene sequences to gain a deeper understanding of the human body (*Bar-Joseph, 2004*). Further, in environmental science and radiology, researchers can use observational data to analyze hydrometeorology (*Shahane, Thomas & Bock, 1977*), climate change (*Gluhovsky & Agee, 2007*), rainfall prediction (*Barrera-Animas et al., 2022*), X-rays, and gamma rays (*Protheroe & Hocking, 1988*). Finally, financial and traffic data are commonly used to predict market fluctuations (*Idrees, Alam & Agarwal, 2019*), stock prices (*Chen, Cheng & Jong Teoh, 2007*; *Spiro et al., 2018*; *Li, Wu & Wang, 2020*), passenger flow (*Ye et al., 2020*), *etc.* Time series is a ubiquitous data type in our daily lives, and the analysis thereof holds great value.

Time series applications are present in every aspect of our lives, computational statistics and data analysis will give us a new perspective and help us gain a deeper understanding of the world.

Motivation

Time series is an important data object, used in an extensive range of research, including classification, prediction, clustering, similarity retrieval, anomaly detection, and noise elimination (*Kalpakis, Gada & Puttagunta, 2001*). The analysis and investigation of its current research applications can provide a comprehensive research review to aid future researchers in understanding the current development state of time series-related algorithms.

To identify the trending topics in current time series research, we further analyze the chosen studies. After removing generic terms like time series, time, analysis of time series, *etc.*, we obtain a co-occurrence map by using literature keyword, shown in Fig. 2.

The font size in the figure is related to the frequency of occurrence of keywords. The larger the font, the higher the frequency of occurrence. There are approximately seven clusters in the figure, representing algorithms and different application domains. Two main research topics are identified, namely, classification and prediction. Because this article focuses on the analysis of time series algorithms, we will present our analysis and conclusions based on the technical development route of classification and prediction algorithms and discuss relevant areas for subsequent research.

Main contribution

The main contributions of this article can be summarized as follows:

- a comprehensive analysis of prevalent topics in the field of time series;
- an investigation into the progress of time series classification and prediction problems in recent years, highlighting several technical development routes that are widely studied in the field, and discussing the improvement and optimization of these algorithms;

Figure 2 **Keyword co-occurrence map of time series literatures (2017–2021).**

- a comparison of the performance of the different algorithms on multiple datasets, concluding with their advantages and disadvantages;
- and finally, an analysis of the challenges and future development tendencies of time series classification and prediction problems.

Methods and materials

The following is the process of our study. First, a literature analysis tool is used to identify current popular research topics. We use VOSViewer to analyze the time series literature through keywords to examine the areas of greatest interest. These topics are classified into 478 categories, and the two research directions with the highest frequency are "classification" and "prediction". Then, the relevant scientific literatures for the identified categories are located. We review related papers on time series classification and prediction and select 87 literatures with high relevance and high citation for analysis. The scientific databases used in the search include Web of Science Core Collection, IEEE Xplore, ACM Digital Library, Springer Link, and ScienceDirect. Finally, according to the literatures,

important technical development routes are extracted, and detailed analysis and summary are carried out.

Structure of this survey

The remainder of this article is organized as follows. 'Related Work' provides an introduction of related work of time series investigation and the comparison of our survey with other traditional surveys and review articles. 'Preliminaries' describes the fundamentals of time series classification and prediction tasks. 'Time Series Classification' elaborates on the development route of time series classification and prediction algorithms, by comparing their performances, analyzing the challenges being faced, and discussing future development trends. Finally, the 'Conclusion' concludes the article.

RELATED WORK

Knowledge discovery in time series is an important direction for dynamic data analysis and processing. The urgent need to predict future data trends based on historical information has attracted widespread attention in many research fields. In the past few decades, many studies have summarized time series research methods from different perspectives. Table 1 summarizes the existing time series surveys and their contributions.

In contrast to the above works, we focus on the development direction of time series technical routes, try to track the most primitive methods of each technical route, study the improvement ideas and improvement strategies of subsequent methods, and compare the advantages and disadvantages of various technical routes and methods. Finally, we provide new ideas for future work.

PRELIMINARIES

Categories of time series

Using data characteristics, time series can be classified into five categories:

1. *Variables*: According to the number of variables, time series can be divided into univariate and multivariate time series. Univariate time series only contains a single variable, while multivariate time series contains multiple variables. For example, *Kadous & Sammut (2005)* use ECG (electrocardiogram) to predict whether patients suffer from heart disease; here ECG can be regarded as a univariate time series. Knape et al. (*Lopez, Steven & Antonio, 2017*) use weather and climate data to analyze its influences on the animal population. These two variables form a multivariate time series.

2. *Continuity*: Time series can be classified as discrete or continuous time series. For example, a gene sequence can be regarded as discrete time series (*Göb, 2006*), while hourly power demand is a continuous time series.

3. *Stability*: Based on the stability of a time sequence, time series can be categorized as a stationary time series or non-stationary time series. The statistical law states that a stationary time series will not change over time. Its sequence diagram intuitively shows random fluctuations around a constant value, within a bounded fluctuation range, with no obvious trends or periodic characteristics. The common periodic function is a

Table 1 Related time series surveys.

Theme	Related surveys	Topic	Key contributions
Prediction	(*De Gooijer & Hyndman, 2006*)	prediction	Review the time series prediction research of the past 25 years.
	(*Mehrmolaei & Keyvanpourr, 2015*)	time series mining; event prediction	Classify and evaluate event prediction methods in time series.
	(*Nyein Naing & Htikemuhammad Yusof, 2015*)	time series prediction; machine learning	Review the time series prediction of machine learning technology in different states spanning ten years.
	(*Mahalakshmi, Sridevi & Rajaram, 2016*)	prediction; time series mining	Provide a detailed survey of various techniques used to predict different types of time series datasets, and discuss various performance evaluation parameters used to evaluate predictive models.
	(*Deb et al., 2017*)	prediction; machine learning; energy prediction	A comprehensive review of existing machine learning techniques used to predict time series energy consumption.
	(*Tealab, 2018*)	prediction; nonlinear time series; neural network	Summarize the research progress of artificial neural network methods in time series prediction models.
	(*Bose & Mali, 2019*)	prediction; fuzzy time series	Summarize and review the contributions in the field of fuzzy time series prediction in the past 25 years.
	(*Hajirahimi & Khashei, 2019*)	prediction; mixed structure	Analyze various hybrid structures used in time series modeling and prediction.
	(*Salles et al., 2019*)	prediction model; non-stationarity; conversion method	Review and analyze the conversion methods of non-stationary time series, and discuss their advantages and limitations on time series prediction problems.
	(*Sezer, Gudelek & Ozbayoglu, 2020*)	prediction; deep learning; finance	Provide research on deep learning in the field of financial time series prediction.
	(*Lim & Zohren, 2021*)	counterfactual prediction; deep neural networks	Survey encoder–decoder designs for time series forecasting and recent developments in hybrid deep learning models.
	(*Liu et al., 2021*)	Intelligent predictors; Hybrid modeling strategies	Analyze various components and combinations in mixed models for time series forecasting.
Classification	(*Radha & Divya, 2017*)	classification; data mining technology	Research multiple time series and classification techniques and investigate various data mining methods for disease prediction.

Table 1 (*continued*)

Theme	Related surveys	Topic	Key contributions
	(*Fawaz et al., 2019*)	deep learning; time series classification	Conduct empirical research on the latest deep neural network architecture for time series classification, and analyze the latest performance of deep learning algorithms for time series classification.
	(*Abanda, Mori & Lozano, 2019*)	classification; distance	Summarize the development of distance-based time series classification methods.
	(*Ali et al., 2019*)	clustering; classification; visualization; visual analysis	Clarify the main concepts of using clustering or classification algorithms in the visual analysis of time series data.
Data mining	(*Chung Fu, 2011*)	data mining; representation; similarity; segmentation; visualization	Comprehensively review the existing research on time series data mining and divide it into research directions such as representation and indexing, similarity measurement, segmentation, visualization, and mining.
	(*Fakhrazari & Vakilzadian, 2017*)	data mining; machine learning	Summarize the existing data mining techniques for time series modeling and analysis and divide the main research directions of time series into three sub-fields: dimensionality reduction (time series representation), similarity measurement, and data mining tasks.
Clustering	(*Rani & Sikka, 2012*)	clustering; data mining; dimensionality reduction; distance measurement	Investigate the clustering of time series in various application fields such as science, engineering, business, finance, economics, health care, and government.
	(*Seyedjamal, Saeed & Wah, 2014*)	time series clustering; subsequence	Review the definition and background of subsequence time series clustering.
	(*Aghabozorgi, Seyed Shirkhorshidi & Ying Wah, 2015*)	clustering; distance measurement; evaluation measures	Reveal the four main components of time series clustering, investigating the improvement trends in the efficiency, quality, and complexity of clustering time series methods over the past decade.

Table 1 (*continued*)

Theme	Related surveys	Topic	Key contributions
	(*Teichgraeber & Brandt, 2022*)	clustering; representative periods	Summarize time series analysis methods used in energy system optimization models.
Similarity measure	(*Chen, Liu & Sun, 2017*)	time series data mining; time series similarity; mining accuracy	Analyze the advantages and disadvantages of current time series similarity measures, and the application of similarity measures in the clustering, classification, and regression of time series data.
	(*Zheng-Xin et al., 2017*)	multivariate time series; data mining; similarity; similarity search	Summarize the existing time series similarity measures, compares different methods of multivariate time series similarity searches, and analyze their advantages and disadvantages.
Deep learning	(*Langkvist, Karlsson & Loutfi,, 2014*)	unsupervised feature learning; deep learning	Review the latest developments in deep learning and unsupervised feature learning for time series problems.
	(*Xu-Dong, 2019*)	deep learning; prediction; classification; anomaly detection	Summarizes the latest deep learning methods for time series prediction, classification, and anomaly detection from the aspects of application, network architecture, and ideas.
	(*Lara-Benítez, Carranza-García & Riquelme, 2021*)	deep learning; forecasting	Evaluate the performance of several deep learning architectures on multiple datasets.
Change detection	(*Zhu, 2017*)	time series change detection	A comprehensive review of the four important aspects of the Landsat time series-based change detection research, including frequency, preprocessing, algorithm, and application.
	(*Namoano et al., 2019*)	online change detection; anomaly detection; time series segmentation	Summarize the main techniques of time series change-point detection, focusing on online methods.
Others	(*Patton, 2012*)	correlation; reasoning; multivariate model; semi parametric estimation	Investigates the estimation, inference methods, and goodness-of-fit test based on copula-based economic and financial time series models, as well as the empirical application of copula in economic and financial time series.

Table 1 (*continued*)

Theme	Related surveys	Topic	Key contributions
	(*Sang, 2013*)	hydrological time series analysis; wavelet transform	Summarizes and reviews the research and application of wavelet transform method in hydrological time series from six aspects.
	(*Nordman & Lahiri, 2014*)	experience likelihood	Summarize the progress of the experience likelihood of time series data.
	(*Tang et al., 2015*)	complexity test	Discuss the complexity testing technology of time series data.
	(*Scotto, Wei & Gouveia, 2015*)	autocorrelation function (ACF); count; sparse operator	Investigate the development of the field of integer-valued time series modeling, and review the literature on the most relevant sparse operators proposed in the analysis of univariate and multivariate integer-valued time series with limited or unlimited support.
	(*Maçaira et al., 2018*)	regression analysis; artificial intelligence; exogenous variables; prediction scheme	A systematic literature review of time series models with explanatory variables.
	(*Papo, 2021*)	irreversibility; time-reversal symmetry	Review and compare important algorithms for testing the irreversibility of time series.

typical stationary time series. However, in real life, non-stationary time series constitute the majority, with examples like wind intensity.

4. *Distribution*: Based on the sequence distribution, a time series can be divided into Gaussian and non-Gaussian time series.

5. *Chaos*: The generation of a chaotic time series is related to its initial conditions, where a change in the initial state of the system may lead to a critical state or inflection point of the interconnected system, significantly impacting on the performance of interconnected system. For example, the action of opening a window or door will affect the power consumption of an air conditioning system (*Kim, 2017*).

Related definitions

To explain time series and its methods in clearer manner, some definitions involving time series are introduced below.

Definition 1. *Univariate time series*: A univariate time series, $s = t_1, t_2, \ldots, t_L$, is an ordered set of length L.

Definition 2. *Multivariate time series*: A multivariate time series, $X = (x_1, x_2, \ldots, x_T)$, is a sequence vector, where each element x_i is a univariate time series, with differing lengths, X has T variables, with the ith variable being x_i.

Definition 3. *Subsequence*: Given a time sequence s with length L, $s_{sub} = s[m, m+n-1]$ is a subsequence with a length $n < L$. The starting point of the subsequence is the position m in s, and the position $m+n-1$ is the end point, represented as $s_{sub} = t_m, \ldots, t_{m+n-1}$, where, $1 \leq m \leq L - n + 1$.

Definition 4. *Similarity degree*: For two time series, b and s, (assuming $|b| \leq |s|$), the similarity degree for them can be computed by $Sim(b, s) = min\{dist(b, s_i)\}$, where s_i is an arbitrary subsequence of s that satisfies the condition $|b| = |s_i|$.

Definition 5. *Shapelet*: A shapelet is a subsequence of time series, s, with the strongest discriminative ability. Specifically, the shapelet can be represented by $p = (b, \delta, c)$, where b, δ, c are the subsequence, threshold, and class label, respectively. If an unknown time series satisfies the condition $Sim(p, s) \leq \delta$, then it can be categorized into class c.

Definition 6. *Euclidean distance*: Euclidean distance is a frequently used distance measurement to determine the degree of similarity of two different time series. For sequences b and c, both with length L, the Euclidean distance can be calculated as $dist_{euclidean} = \sqrt{\sum_{i=1}^{L}(b_i - c_i)^2}$.

Definition 7. *Dynamic time warping (DTW)*: DTW is another widely used distance measurement method. Compared with Euclidean distance, it can compute the minimum distance between two sequences with different lengths. For its wide application, the principle will not be explained here, but the calculation is given as $dist_{DTW} = DTW(s, b)$.

Basic algorithms

In time series classification and prediction tasks, the most basic and widely used algorithms are 1NN-DTW (1 nearest neighbor dynamic time warping) and autoregressive (AR) and moving average (MA) models.

1NN-DTW

The 1NN-DTW model uses DTW as distance measurement, and the simple but effective algorithm 1NN to find the nearest training sample of the current instance and assigns the same class label to the instance as the nearest training sample. This model does not require training of parameters and has high accuracy. The following pseudocode describes the procedure of 1NN-DTW.

Algorithm 1 1NN-DTW

Require: T: labeled time series dataset, the number of samples is N
Ensure: acc: average 1NN classification accuracy
 1: $Num = 0$
 2: **for** each instance s_i of T **do**
 3:     $distance = DTW(s_i, T - s_i)$;
 4:     assign the closest instance label y^{pred} of T to s_i;
 5:     **if** $y^{pred} == y^{s_i}$ **then**
 6:         $Num = Num + 1$;
 7:     **end if**
 8: **end for**
 9: $acc = \frac{Num}{N}$;

AR and MA

- AR model

The model is represented as $X_t = \sum_{j=1}^{p} a_j X_{t-j} + \varepsilon_t$ and is called the p-order AR model, denoted as $AR(p)$, where a time series value can be expressed as a linear function of its previous value, X_t, and an impact value, ε_t. This model is a dynamic model that is different from the static multiple regression model.

- MA model

The model is represented as $X_t = \varepsilon_t + \sum_{j=1}^{q} b_j \varepsilon_{t-j}$ and is called the q-order MA model, denoted as $MA(q)$. The time series value, X_t, is the linear combination of the present and past error or shock value, ε_t.

TIME SERIES CLASSIFICATION

Unlike traditional classification tasks, the order of the time series variable is related to the input object, which makes time series classification a more challenging problem. Based on data label availability, current time series classification research mainly focuses on supervised and semi-supervised learning. Usually, supervised learning methods with labeled information show better performances. However, in real life there is a tremendous amount of unlabeled data. Therefore, some semi-supervised methods have been proposed to address this situation by constructing models using limited labeled data and a large amount of unlabeled data. In addition, some specific application scenarios have new requirements for time series classification tasks, for example, the early diagnosis of a disease, which results in a better prognosis. Early classification is used in these situations, and its goal is to classify data as soon as possible with a certain accuracy rate. This is an important extension of time series classification. This section introduces the development route of time series classification technology, analyzes the current difficulties and challenges, and mentions some expected future trends.

Technology developments

Based on the literature reviewed, we discover three development routes: supervised time series classification, semi-supervised time series classification, and early classification, which is a critical extension of the time series classification task. Fig. 3 lists the algorithms of different technology development routes.

Supervised learning

In early time series classification methods, the work mainly focus on the distance-based algorithm (*Ding et al., 2008*). The most prominent one being 1NN-DTW, which has demonstrated excellent performance in multiple tasks and datasets (*Ding et al., 2008*), and was once considered as an insurmountable method in time series classification (*Xi et al., 2006*; *Ye & Keogh, 2009*; *Rakthanmanon & Keogh, 2013*). With the deepening of related research, some algorithms with better performance, such as Rocket (*Dempster, Petitjean & Webb, 2019*), have achieved better results than 1NNDTW on multiple data sets. Even so, 1NNDTW is worthy of analysis and academic attention. The 1NN-DTW uses 1NN as a classifier, DTW as distance measurement criteria, and assigns the nearest training
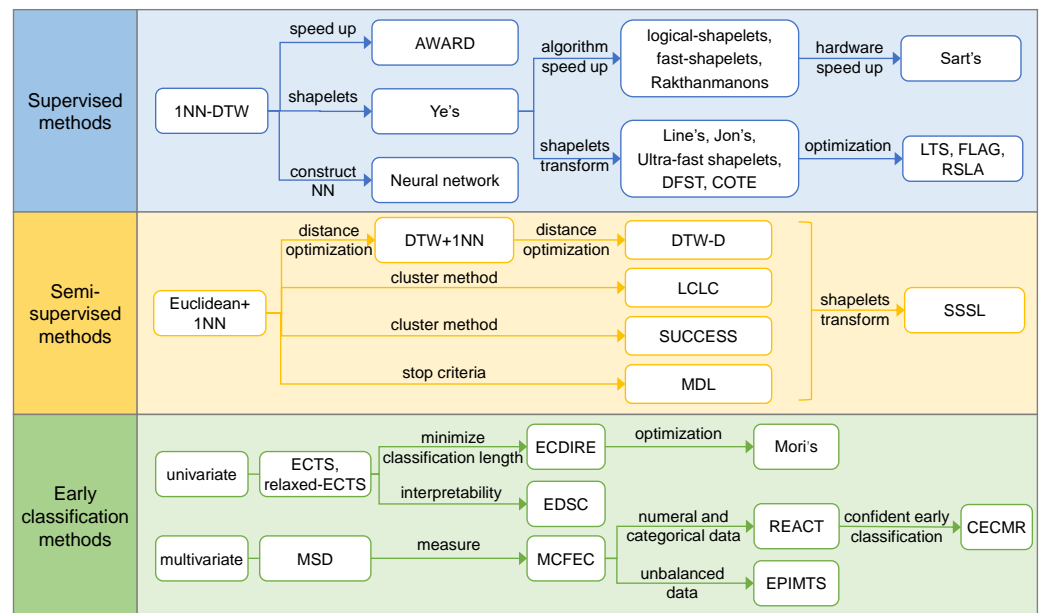
Figure 3 Technology development routes.

instance class label to a testing instance. This algorithm is simple, and has a high accuracy. In practice, training on the optimal hyperparameter settings, such as warping windows, is required to obtain better performance (*Dau et al., 2018*). However, during the classification stage, the class label of every testing instance needs to be computed by working through the entire training dataset, affording it high time complexity. The optimization of 1NN-DTW mainly concentrated on reducing classification time, by using one of three methods.

- *Speed up*

The idea of this type of algorithm is that the effectiveness can be improved by reducing the dataset size and accelerating the computation of DTW. Through numerosity reduction and dynamic adjustment of the DTW warping window size (*Xi et al., 2006*), 1NN-DTW can be sped up while guaranteeing accuracy.

- *Shapelets*

Geurts (2001) propose that a time series can be represented by its local pattern. Based on this idea, *Ye & Keogh (2009)* formally propose the concept of shapelets. The most important idea of shapelets is to extract the most discriminative subsequence from the whole sequence, and then making a classification by constructing a decision tree.

The advantages of the shapelet-based method are that it has strong interpretability, robustness, and low classification time complexity. Although it can be accelerated through early abandon and entropy pruning methods, the search space and time complexity of shapelets are still not negligible. Therefore, some acceleration strategies such as precomputing of reusable distance and allowable pruning (*Mueen, Keogh & Young, 2011*), discrete representation of subsequence (*Rakthanmanon & Keogh, 2013*), early abandoning Z-normalization, reordering early abandoning, reversing the query/data role, and cascading

lower bounds (*Rakthanmanon et al., 2012*), are applied in the search of shapelets. In addition, some studies use shapelet transform to construct a new dataset from the original dataset, expecting reduced training time while retaining model interpretability and further improving accuracy (*Lines et al., 2012*; *Hills et al., 2014*). Shapelet transform separates the search procedure of shapelets and classifier construction (using the distance between shapelets and the original sequence as a new dataset), and this makes the selection of the classifier flexible.

Since the advent of shapelet transform, subsequent research has shifted focus to identify more effective ways of finding shapelets (*Wistuba, Grabocka & Schmidt-Thieme, 2015*; *Baldán & Bentez, 2018*). In contrast to constantly searching for shapelets in existing sequences, some algorithms (*Grabocka et al., 2014*; *Bagnall et al., 2015*; *Hou, Kwok & Zurada, 2016*; *Zhao, Pan & Tao, 2020*) believe that shapelets can be learned, and this changes the shapelet searching process into a mathematical optimization task, which can improve the performance of the model. However, some methods consider the performance of acceleration technology to be close to the upper bound, so other solutions must be considered, such as using multiple GPUs and FPGAs to accelerate the DTW subsequence search process (*Sart et al., 2010*).

- ● *Constuct of a neural network*

This type of algorithm is a feature-based method, and its main idea is to train the classifier in advance. *Iwana, Frinken & Uchida (2020)* embed DTW into a neural network as a kernel function. In this way, the neural network can solve the problem of time series sequence recognition, such as time distortion and variable pattern length, in feedforward architecture. There have been many studies devoted to applying deep learning models to time series classification (*Zheng et al., 2014*), and *Fawaz et al. (2019)* provide a detailed introduction and summary.

Using the results from previous studies, we compare the accuracy of various methods (as shown in Table 2) with multiple public datasets which are widely used in this field (*Ding et al., 2008*; *Rakthanmanon & Keogh, 2013*; *Lines et al., 2012*). The performance of the shapelets learning method (LTS, FLAG, RSLA) is superior. According to the different principles used in the methods, we divide the algorithms into five categories: 1NN-DTW, shapelets, shapelets transform, shapelets learning, and neural network. In addition, the advantages and disadvantages of 1NN-DTW, shapelets, shapelets transform, and shapelets learning are compared in Table 3.

1NN-DTW is the simplest, high performing method that needs no training and can correctly classify samples. However, its biggest problem is long classification times, especially for large training datasets, which makes it unsuitable for certain applications. The shapelets-based method reduces the sequence length, and thus, has a faster classification time, and achieves high interpretability and robustness. However, shapelets are discriminative features that require significant effort to find, and for large sequence lengths, the search space increases drastically. The shapelet transform method makes the choice of classifier more flexible, but it still retains the long search time problem. The shapelet learning method learns the shapelets instead of searching for them through

Table 2 Comparison of the accuracy of supervised time series classification.

Category	1NN-DTW	Shapelets	Shapelets transform			Shapelets learning			
	1NN-DTW	Fast shapelets	Shapelet transform	COTE	LTS	FLAG	RSLA-LS	RSLA-LZ	
	(Ding et al., 2008)	(Rakthanmanon & Keogh, 2013)	(Lines et al., 2012)	(Bagnall et al., 2015)	(Grabocka et al., 2014)	(Hou, Kwok & Zurada, 2016)	(Zhao, Pan & Tao, 2020)	(Zhao, Pan & Tao, 2020)	
Adiac	60.0	54.9	29.2	76.9	49.7	74.2	**75.4**	73.9	
Beef	63.3	56.7	50.0	80.0	83.3	80.0	83.3	**86.7**	
Chlorine	64.8	59.1	58.8	68.6	59.4	78.0	75.0	**81.4**	
Coffee	**100.0**	96.4	96.4	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	
Diatom	96.4	87.9	72.2	89.2	96.7	96.4	96.7	**97.7**	
DP_Little	50.3	57.8	65.4	–	**71.7**	65.7	69.1	69.8	
DP_Middle	54.1	59.2	70.5	–	73.5	72.9	72.6	**73.8**	
DP_Thumb	53.0	59.1	58.1	65.4	**75.7**	72.4	70.7	75.0	
ECGFiveDays	78.7	99.5	77.5	99.9	**100.0**	92.0	**100.0**	**100.0**	
FaceFour	82.9	92.0	84.1	71.6	95.4	90.9	92.0	**95.5**	
Gun_Point	94.0	94.0	89.3	93.3	**100.0**	96.7	96.7	99.3	
ItalyPower	95.2	90.5	89.2	96.2	95.9	94.6	96.5	**96.8**	
Lighting7	73.9	63.0	49.3	61.6	78.1	76.7	75.3	**79.5**	
MedicalImages	**74.3**	60.5	48.8	67.1	67.8	72.4	71.4	73.4	
MoteStrain	86.8	79.8	82.5	84.0	85.1	88.8	**89.5**	**89.5**	
MP_Little	55.2	62.1	66.4	–	**73.9**	71.8	73.6	73.6	
MP_Middle	55.2	61.7	71.0	–	77.3	76.6	74.7	**78.3**	
Otoliths	59.3	60.9	–	60.9	67.2	64.1	**73.4**	71.9	
PP_Little	55.2	48.7	59.6	–	**72.7**	68.5	71.6	70.5	
PP_Middle	50.0	56.8	61.4	–	74.9	74.0	72.7	**75.2**	
PP_Thumb	51.2	58.9	60.8	–	70.1	68.4	69.8	**70.7**	
Sony	73.2	68.5	–	87.7	85.3	92.8	93.2	**95.3**	
Symbols	94.1	93.6	78.0	**94.7**	93.9	87.5	91.3	92.3	
SyntheticC	99.3	93.6	94.3	81.0	**99.7**	**99.7**	**99.7**	99.0	
Trace	**100.0**	**100.0**	98.0	**100.0**	**100.0**	99.0	98.0	**100.0**	
TwoLeadECG	89.3	94.6	85.0	91.6	**99.7**	99.0	99.3	99.3	

Notes.
A dash (-) indicates that there is no data available. The bold values represent the highest accuracy for each category.

training data, so the learned shapelets have higher robustness compared to the searched one. The disadvantage of this type of method is the long training time required.

Semi-supervised learning

Semi-supervised learning methods construct classifiers using a small amount of labeled data and a large amount of unlabeled data. One of the most frequently used methods is self-learning: it utilizes a small amount of labeled data to assign class labels to a large unlabeled dataset.

Wei & Keogh (2006) propose extending training data by 1NN, if the distance between the labeled data and unlabeled data is close enough, then add the unlabeled data into the training set. This is a simple and basic semi-supervised learning approach for time series classification. Based on this, the subsequent advancements can be divided into three categories.

Table 3 Comparison of supervised time series classification.

Category	Methods	Advantages	Disadvantages
1NN-DTW	1NN-DTW (*Ding et al., 2008*) AWARD (*Xi et al., 2006*)	Simple, no training needed	High time complexity of classification
Shapelets	Ye's (*Ye & Keogh, 2009*), logical-shapelets (*Mueen, Keogh & Young, 2011*), fast-shapelets (*Rakthanmanon & Keogh, 2013*), Rakthanmanon's (*Rakthanmanon et al., 2012*), Sart's (*Sart et al., 2010*)	High interpretability and robustness, low classification time complexity	High time complexity of shapelets searching procedure, and for large length sequences, the time cost becomes unacceptable
Shapelet transform	Line's (*Lines et al., 2012*), Jon's (*Hills et al., 2014*), Ultra-fast shapelets (*Wistuba, Grabocka & Schmidt-Thieme, 2015*), DFST (*Baldán & Bentez, 2018*) COTE (*Bagnall et al., 2015*)	High accuracy and flexible	Long shapelets search time
Shapelet learning	LTS (*Grabocka et al., 2014*), FLAG (*Hou, Kwok & Zurada, 2016*), RSLA (*Zhao, Pan & Tao, 2020*)	High robustness, interpretability, discriminativeness	Long training time

• *Distance*

Wei & Keogh (2006) use Euclidean distance as a similarity measurement; because DTW is a more effective distance in time series classification, it can be used to improve model performance (*Chen et al., 2013*). However, the ratio of DTW and Euclidean distance is proposed to be the proper distance measurement, making the algorithm more suitable for smaller data sizes and diverse negative samples. This is based on two assumptions: first, negative samples are diverse, and the negative samples may have a closer distance with positive samples; second, compared with Euclidean distance, DTW makes the inter-distance of positive samples closer.

• *Label approach*

Other than optimizing the distance function, changing the method of adding testing data into the training dataset can also improve classification results. One possible way is to cluster negative samples. Because a robust classifier needs to be constructed using limited, labeled, positive data, partitioning the unlabeled dataset into smaller local clusters, and identifying the local clusters' common principal features for classification can make the algorithm more reliable and productive (*Nguyen, Li & Ng, 2011*). Hierarchical clustering is also an effective cluster method (*Marussy & Buza, 2013*); first, it clusters all sequences into smaller clusters, and then uses seeds to assign labels to unlabeled data.

• *Stopping criterion*

If a stopping criterion is too conservative (or too liberal), it is doomed to produce many false negatives (or false positives) (*Begum et al., 2013*). Therefore, it is important to propose a proper stopping criterion to avoid adding negative samples into the positive sample set. *Begum et al. (2013)* propose a parameter-free algorithm for finding a stopping criterion using the minimum description length (MDL) technique. The algorithm is stopped when the MDL becomes large, improving the classification results by optimizing the stopping criterion (*Rodriguez, Alonso & Bostrom, 2001*).

Table 4 Comparison of the accuracy of semi-supervised classification methods.

Datasets	Class number	Wei (*Wei & Keogh, 2006*)	DTW-D (*Chen et al., 2013*)	SUCCESS (*Marussy & Buza, 2013*)	Xu (*Xu & Funaya, 2015*)	SSSL (*Wang et al., 2019*)
Coffee	2	57.1	60.1	63.2	58.8	79.2
CBF	3	99.5	83.3	99.7	92.1	100.0
ECG	2	76.3	95.3	77.5	81.9	79.3
Face four	4	81.8	78.2	80.0	83.3	85.1
Gun point	2	92.5	71.1	95.5	72.9	82.4
ItalyPow.Dem	2	93.4	66.4	92.4	77.2	94.1
Lighting2	2	65.8	64.1	68.3	69.8	81.3
Linghting7	7	46.4	50.3	47.1	51.1	79.6
OSU leaf	6	46.0	70.1	53.4	64.2	83.5
Trace	4	95.0	80.1	100.0	78.8	100.0
WordsSyn	25	59.0	86.3	61.8	63.9	87.5
OliveOil	4	63.3	73.2	61.7	63.9	77.6
StarLight Curves	3	86.0	74.3	80.0	75.5	87.2

The accuracy of different semi-supervised methods is compared in Table 4, as collected from various studies. The overall performance of the SSSL method is the best, which shows that the method of learning shapelets through optimization algorithms is still effective in semi-supervised learning. While shapelets improve accuracy, they also improve the interpretability of the algorithm, again highlighting the importance and usefulness of shapelets.

Early classification

The main goal of early classification is to assign class labels as early as possible while guaranteeing a certain percentage of accuracy. It has great importance in time sensitive applications, such as the diagnosis of heart disease, as early diagnosis improves prognosis. In practical applications, due to an unclear description of the issues to be solved, the early classification of time series may cause false positives in practical applications, and the cost of false positives is very high. To solve this problem, *Wu, Der & Keogh (2021)* propose that the definition of early classification of time series should be clearly defined first, and it is also very important to obtain real-world publicly available datasets. According to the data type, there are two technology development routes.

- **Univariable**

Rodriguez, Alonso & Bostrom (2001) segment a time series into intervals and then describe these intervals using relative predicates and region-based predicates. It is the first literature to mention the term early classification of time series. Although it achieves early classification by using sub-information, it does not consider ways to choose the shortest prefix to provide reliable classification results. ECTS (*Xing, Pei & Yu, 2009*) obtains the shortest prediction length through training, and it uses the sequence prefix to classify data under the condition of guaranteed accuracy. ECTS achieves a shorter prefix, higher accuracy, and higher effectiveness by using an accelerating algorithm. Further, *Mori et al. (2016)* calculate the shortest classification length for each class, while *Mori et al. (2018)*

Table 5 Comparison of univariable accuracy in early classification.

Methods	Datasets						
	Wafer	Gun Point	Two patterns	ECG	Synthetic control	OliveOil	CBF
ECTS (*Xing, Pei & Yu, 2009*)	99.08	86.67	86.48	89.00	89.00	90.00	85.20
RelaxedECTS (*Xing, Pei & Yu, 2012*)	99.08	86.67	86.35	89.00	88.30	90.00	85.20
ECDIRE (*Mori et al., 2016*)	97.00	87.00	87.00	91.00	96.00	40.00	89.00
EDSC (*Xing et al., 2011*)	99.00	94.00	80.00	85.00	89.00	60.00	84.00

change this task into a mathematical optimization problem, using the accuracy and earliness as mutual optimization goals.

The above methods lack interpretability, which is useful in determining the factor affecting an object. EDSC (*Xing et al., 2011*) introduces shapelets and proposes local-shapelets, using kernel density estimation or Chebyshev inequality to find the threshold value of each shapelet, and then selecting the best shapelet for classification.

● *Multivariable*

MSD (*Ghalwash & Obradovic, 2012*) extends the EDSC algorithm to suit a multivariable situation. It uses information gain to evaluate the goodness of shapelets, adds shapelet pruning, and abandons shapelets that has no ability to correctly classify data. This method has three disadvantages: first, it handles multivariable data in a fixed window, even though, different variables could have different shapelet positions; second, it cannot process variables with different lengths; and third, there is no connection between multiple variables.

To solve these problems, *He et al. (2015)* propose learning a shapelet for each variable independently, and constructing a classifier that can use multiple shapelets to classify data. Moreover, it substitutes information gain with a new measurement (F-measure). This method can solve the inter-class imbalance problem (a class containing multiple small classes, or consisting of multiple concepts) to a certain degree through inter-class clustering. *Lin et al. (2015)* further extend the input variables of the algorithm from continuous numerical sequences to characterized discrete sequences. *He et al. (2019)* use downsampling technology to solve the intra-class imbalance problem, and a clustering method to deal with the inter-class imbalance problem, which further expands the applicability of the algorithm.

In contrast, *He, Zhao & Xia (2020)* mainly focus on the identification of multivariable class labels as early as possible and ensures the classification accuracy higher than the probability of true label. Tables 5 and 6 compare the accuracy of some univariate early classification algorithms and multivariate early classification algorithms, respectively.

While most of the univariable classification algorithms achieve good results (above 85%), the accuracy of multivariable algorithms do not reach that high (except EPIMTS). This can be attributed to the fact that it is difficult to consider multiple variables simultaneously and extracting the interconnection between them correctly. EPIMTS uses an ensemble method to combine these two important factors into the algorithm, allowing it to achieve the best performance.

Table 6 Comparison of multivariable accuracy in early classification.

Methods	Datasets			
	Syn1	Syn2	Wafer	ECG
Class number	2	3	2	2
Variable number	3	4	6	2
MSD (*Ghalwash & Obradovic, 2012*)	0.74	0.34	0.74	0.74
MCFEC-QBC (*He et al., 2015*)	0.99	0.77	0.9	0.77
MCFEC-Rule (*He et al., 2015*)	0.98	0.74	0.97	0.78
EPIMTS (*He et al., 2019*)	0.98	0.99	0.96	0.84

Challenges and future trends

This section discusses the different technology development routes in time series classification. Mainly, the research covers both traditional supervised learning methods and semi-supervised learning methods. In particular, an important extension—early classification—is proposed for specific application situations.

Although the existing work has achieved good results in time classification tasks, there are still some problems. In real life, the amount of unlabeled data exceeds that of labeled data, and its sources are more abundant. Although supervised learning yields better classification results, labeling data is expensive and time consuming. In some fields such as medical and satellite data, experts are required to label the data, making the acquisition of labeled data even more difficult. Therefore, research on semi-supervised or unsupervised methods has great value. However, according to the research reviewed for this article, very few recent studies focus on semi-supervised learning methods and unsupervised learning methods for time series classification (*Wei & Keogh, 2006*; *Chen et al., 2013*; *Nguyen, Li & Ng, 2011*). Managing large amounts of unlabeled data for classification tasks is a tremendous challenge we face.

TIME SERIES PREDICTION

Although time series prediction methods have experienced a long period of development, the rapid increase in data scale has brought severe challenges to traditional time series prediction methods, and has also seriously affected the efficiency of prediction methods. Time series prediction methods have gradually developed from simple linear regression models and nonlinear regression models based on traditional statistics to machine learning methods represented by neural networks and support vector machines. At the same time, researchers have also proposed other prediction methods for time series with different characteristics based on different theoretical foundations. Fuzzy cognitive map can deal with data uncertainty and maintain a high level of interpretability. To solve the problem of insufficient labeled data for some practical applications, transfer learning methods can be used. Two future research avenues are clear; first, dealing with rapid increase in the scale of time series data; second, choosing the most suitable model for a specific problem.
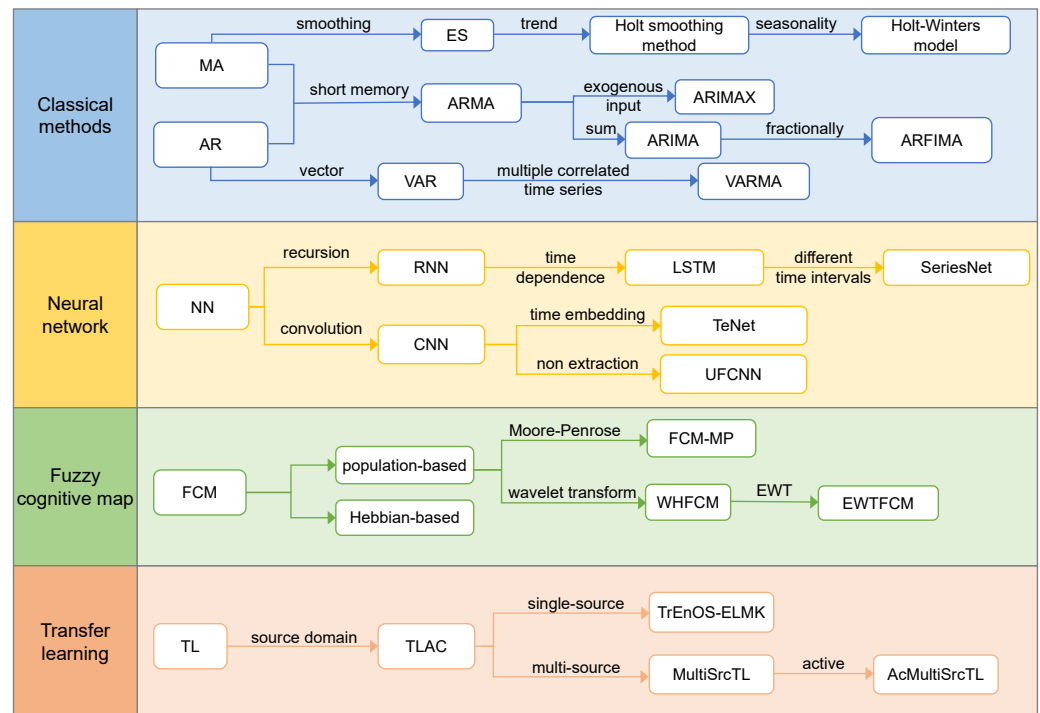
Figure 4 Technology development routes.

Technology developments

According to the reviewed literature, we have defined four technical development routes, namely, the classic algorithm, neural network, fuzzy cognitive map, and transfer learning. Figure 4 lists the development directions of the different technical routes and their resulting algorithms.

Classical methods

The traditional time series prediction methods are mainly used to solve the model parameters on the basis of determining the time series parameter model and using the solved model to complete the prediction work, mainly from the perspective of a stationary series, non-stationary series, or multivariate time series.

● **Stationary series**

Russian astronomer Slutzky create and propose the moving average (MA) model (*Slutzky, 1937*), and British statistician G.U. Yule propose the autoregressive(AR) model (*Yule, 1927*) when studying sunspots. The AR model is a representation of a random process, and its output variable depends linearly on its previous value and random conditions. The purpose of the AR model is to minimize the square error between the predicted results and the actual results. Box and Jenkins propose a short memory model called autoregressive moving average (ARMA) model (*Box & Jenkins, 1970*). The ARMA model provide a general framework for predicting stationary observation time series data. However, it is not suitable for non-stationary time series data, and only one time series can be modeled at a time.

● *Non-stationary series*

Non-stationary time series comprise four trends: long-term trend, cyclic trend, seasonal trend, and irregular trend. Box and Jenkins propose the autoregressive integrated moving average (ARIMA) model for non-stationary short memory data with obvious trends (*Box & Jenkins, 1970*). The ARIMA model has become one of the most widely used linear models in time series prediction. This model uses historical data of univariate time series to analyze its own trends and predict future cycles, but the ARIMA model cannot easily capture non-linear patterns. One or more time differentiation steps in ARIMA keep the time series data unchanged. Differentiation operations usually amplify high-frequency noise in time series data, thereby affecting the accuracy of prediction. When modeling time series with long memory dependence, a common alternative is autoregressive partial integration moving average (ARFIMA). The model is based on ARIMA and allows the difference parameters to be set to non-integer values. On the basis of the ARIMA model, the autoregressive integrated moving average (ARIMAX) model is obtained by adding exogenous input (*Wangdi et al., 2010*).

The exponential smoothing (ES) (*Gardner, 1985*) model is a time series data smoothing technique that uses past data points in a time window to smooth current data points. In contrast to the traditional MA model, the ES model uses an exponential function to assign more weight to the nearest data point, which is beneficial for processing non-stationary time series data, and is aimed at series without trend and seasonality. The Holt smoothing method (*Holt, 2004*; *Winters, 1960*), also called double exponential smoothing, is an extension of ES designed for time series with a trend but no seasonality. *Chatfield (1978)* propose the Holt-Winters model, which uses three smoothing steps to predict time series data. The three smoothing steps are used for level, trend, and seasonality, and are also called three exponential smoothing. The Holt-Winters model can be used for univariate time series prediction of seasonal data.

● *Multivariate time series*

The vector autoregressive (VAR) (*Mizon, 1991*) model is a natural extension of the univariate AR model over dynamic multivariate time series, providing predictions superior to univariate time series models and theory-based fine simultaneous equation models. The vector autoregressive moving average (VARMA) (*Athanasopoulos & Vahid, 2008*) model allows several related time series to be modeled together, considering the cross-correlation and internal correlation of the series. The VARMA model fully considers the influence of each sequence on another sequence, thereby improving the prediction accuracy. This makes the predictions generated by the VARMA model more reliable for decision-making.

Traditional research methods mostly use statistical models to study the evolution of time data. For decades, linear statistical methods have dominated the prediction. Although linear models have many advantages in implementation and interpretation, they have serious limitations in capturing the nonlinear relationship in the data, which is common in many complex real-world problems.

Neural Network

An artificial neural network (ANN) is a flexible computing framework and general approximator that can be applied to various time series prediction problems with high accuracy. The main advantage of a neural network is its flexible nonlinear modeling ability, without the need to specify a specific model form. The popularity of ANN stems from being a generalized nonlinear prediction model. Since the advent of the simplest ANN, the ideas of recursion, nonlinear regression, and convolution continues to develop. According to the characteristics of real data, the linear and nonlinear models can be combined to construct a hybrid model to achieve better performance.

- ● *Recursion*

Connor & Atlas (1991) apply a recurrent neural network (RNN) to time series prediction, using the historical information of time series to predict future results. *Hochreiter & Schmidhuber (1996)* proposes an improved RNN called long short-term memory (LSTM), which solves the problem of the vanishing gradient by introducing additional units that can store data indefinitely, and has shown success in single-step time series analysis. LSTM is able to address sequences of varying length and capture long-term dependencies without the same problems as traditional RNN architectures (*Wilson et al., 2018*). LSTM has gradually become a popular solution for learning the long-term time-dependent characteristics of original time series data, and can use a fixed-size time window to solve many time series tasks that feedforward networks cannot solve.

- ● *Convolution*

Convolutional neural network (CNN) is different from RNN, which strictly uses sequential learning processes. The latter processes one data point each time to generate data representations, while the former use nonlinear filters based on multiple dataset learning representation. In each step, a filter is used to extract features from a subset of local data, so that the representation is a set of extracted features. *Liu et al. (2015)* use a CNN combined with time-domain embedding to predict periodic time series values; a novel model called a time-embedding enhanced convolutional neural network (TeNet), to learn the repeated occurrences in periodic time series structural elements (called abstract fragments) that have not been hidden to predict future changes. *Mittelman (2015)* propose a non-decimated full convolutional neural network (UFCNN) to deal with time series problems. UFCNN has no gradient disappearing and gradient explosion problems, so it is easier to train. It can be implemented more efficiently because it only involves convolution operations instead of the recursion used by RNN and LSTM.

- ● *Hybrid model*

Modeling real-world time series is a particularly difficult task because they usually consist of a combination of both linear and nonlinear patterns. In view of the limitations of linear and nonlinear models, hybrid models have been proposed in some studies to improve the quality of prediction. The ARIMA model, ANN model (*Peter & Zhang, 2003*; *Khashei & Bijari, 2010*; *Babu & Reddy, 2014*), and multi-layer perceptron(MLP) (*de O. Santos Jnior, de Oliveira & de Mattos Neto, 2019*) are combined to construct a hybrid model, which has been proven by experiments to achieve better performance than a single model.

Table 7 Performance comparison of different methods.

Methods	S&P 500 Index			Shanghai Composite Index			Hangzhou Temperature		
	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2
ANN	24.22	20.21	0.965	66.25	39.35	0.975	2.95	2.14	0.895
UFCNN (*Mittelman, 2015*)	24.36	19.84	0.965	93.06	57.77	0.950	**2.64**	**1.97**	**0.907**
LSTM	19.04	14.42	0.978	**63.84**	**38.05**	**0.976**	2.86	2.09	0.901
SeriesNet (*Shen et al., 2020*)	**17.32**	**13.15**	**0.982**	63.94	38.37	**0.976**	2.82	2.06	0.903

Notes.
The data are obtained from reference (*Shen et al., 2020*).

Table 8 Performance comparison of different methods.

Method category	Methods	Advantages	Disadvantages
Classical method	AR, MA, ARMA, ARIMA	Good at linear problems	Cannot handle nonlinear problems well
Traditional machine learning	SVM, LS-SVM (*Suykens & Johan, 2002*)	Able to solve complex time series data	Cannot handle nonlinear problems well
NN	ANN, BPNN, DE-BPNN (*Wang, Zeng & Chen, 2015*)	Able to handle nonlinear problems	Long-term dependence cannot be effectively preserved
LSTM	LSTM	Capable of capturing long-term dependence, structure is conducive to dealing with sequence problems	Facing the problem of gradient disappearance or gradient explosion, and it is difficult to train
CNN	CNN, UFCNN (*Mittelman, 2015*)	Efficient	Difficult to capture long-term dependence
Hybrid model	ARIMA-ANN (*Peter & Zhang, 2003*; *Babu & Reddy, 2014*), ARIMA-SVM (*Pai & Lin, 2005*; *Oliveira & Ludermir, 2014*), ARIMA-NN (*Khashei & Bijari, 2010*), ARIMA-MLP-SVR (*de O. Santos Jnior, de Oliveira & de Mattos Neto, 2019*), SeriesNet (*Shen et al., 2020*)	Better performance	High complexity

Shen et al. (2020) propose SeriesNet, using LSTM and extended random convolution to extract features with different time intervals from the time series, and combining them. This can make full use of the characteristics of the time series and help improve prediction accuracy. Compared with other models, the SeriesNet model has the best prediction accuracy in nonlinear and non-stationary datasets. In the non-stationary datasets, the error of SeriesNet decreases slowly as the size of the sliding window increases.

Table 7 compares the root-mean-square error (RMSE), the mean absolute error (MAE) and the coefficient of determination (R^2) of multiple methods. We summarize the advantages and disadvantages of different methods, and the results are presented in Table 8. The hybrid model has a stronger advantage when dealing with nonlinear and non-stationary data.

Fuzzy cognitive map

Fuzzy cognitive map (FCM) is a dynamic system quantitative modeling and simulation method proposed by *Kosko (1986)*. It is a simple and powerful tool that is very useful in dynamic system simulation and analysis. FCM can be useful in time series prediction tasks that do not need to deal with exact numbers but only need approximate results (*Felix et al., 2019*). This method combines the characteristics of fuzzy logic and neural networks, which can effectively model the states of the system. It can simultaneously deal with the uncertainty of data and maintain a high level of interpretability. It has been demonstrated that FCM can be applied to predict time series with univariate (*Lu, Yang & Liu, 2014*) and multivariate (*Froelich et al., 2012*; *Papageorgiou & Froelich, 2012a*; *Papageorgiou & Froelich, 2012b*; *Stach et al., 2005*) variables.

The existing algorithms applied to train FCM belong to two main groups, population-based and Hebbian-based methods. Population-based algorithms include particle swarm optimization (PSO) (*Homenda, Jastrzebska & Pedrycz, 2015*; *Salmeron et al., 2017*), genetic algorithm(GA) (*Yesil et al., 2013*), memetic algorithms(*Salmeron, Ruiz-Celma & Mena, 2016*), artificial bee colony(ABC) (*Yesil et al., 2013*), and modified asexual reproduction optimization (*Salmeron et al., 2019*). Hebbian-based learning algorithms are seldom used for time series prediction because of their poor generalization ability.

FCM in the time series prediction domain is mostly composed of two parts, establishing the structure and learning the weight matrix. To facilitate an efficient extraction of concepts, FCM framework is constructed by using fuzzy c-means algorithm (*Lu et al., 2014*). When applying standard FCM to time series prediction, most of the literature (*Lu, Yang & Liu, 2014*; *Poczeta & Yastrebov, 2014*; *Papageorgiou, Poczeta & Laspidou, 2015*; *Poczeta, Yastrebov & Papageorgiou, 2015*) assumes that the weights of FCM are adjusted during the training phase and do not change with time when used for prediction. To improve the accuracy of prediction and reduce training time, some studies proposed pseudo-inverse learning and wavelet transform.

- *Pseudo-inverse learning*

Vanhoenshoven et al. (2020) propose a new FCM learning algorithm based on the Moore Penrose inverse (FCM-MP). The unique feature of this learning method is that for the pseudo-inverse learning of the FCM weight matrix, each iteration step calculates a different set of weights. In this way, different time-varying data segments will affect the weight, and the weight will change from one iteration to the next. This algorithm improves the accuracy of prediction, does not require laborious parameter adjustments, and reduces the processing time required for training the FCM.

- *Wavelet transform*

Although fuzzy cluster analysis has strong time series modeling capabilities, prediction methods based on fuzzy cluster analysis cannot handle non-stationary time series, and evolutionary learning methods are not suitable for large-scale time series. To overcome these two limitations, *Yang & Liu (2018)* propose wavelet high-order fuzzy cognitive map (WHFCM), which uses wavelet transform instead of fuzzy time series, and uses ridge regression to train. Further, empirical wavelet transform (EWT) is superior to discrete wavelet transform in time series prediction, because empirical wavelet transform is a

data-driven signal decomposition algorithm. *Gao, Du & Yuen (2020)* propose the mixed time series prediction model based on EWT and FCM. EWT is applied to decompose the original time series into different levels to capture information of different frequencies, and to train high-order fuzzy cognitive maps to model the relationship between all generated subsequences and the original time series.

FCM has been successfully used to model and predict stationary time series. However, it is still challenging to deal with large-scale non-stationary time series with time trends and rapid changes over time. The main advantage of the FCM-based model is the human-centered knowledge representation interface. Therefore, in terms of accuracy, fuzzy admissible mapping time series modeling may not exceed the classical methods that have been studied, but FCM provides superior practical characteristics.

Transfer learning

Time series data usually change over time. Hence, samples collected over a long period of time are usually significantly different from each other. As such, it is generally not recommended to directly apply old data to the prediction process. For time series prediction problems, we hope to train an effective model with only a small number of fresh samples and relatively rich, old data. Therefore, to solve the problem of insufficient labeled data available in some practical applications, transfer learning methods can be used. Transfer learning is the reusing and transferring of knowledge in one field to other different but related fields. Its basic idea is to utilize the data or information of related source tasks to assist in modeling for the target task. Traditional machine learning techniques try to learn each task from scratch, while transfer learning techniques try to transfer the knowledge from some previous tasks to a target task when the latter has less high-quality training data (*Pan & Yang, 2010*).

Xiao, He & Wang (2012) propose a transfer learning-based analog complexing model (TLAC). First, it transfers related time series from the source domain to assist in modeling the target time series using the transfer learning technique. *Ye & Dai (2018)* propose a hybrid algorithm based on transfer learning, combining online sequential extreme learning machine with kernel (OS-ELMK) and integrated learning (TrEnOS-ELMK). With TrEnOS-ELMK, a single-source transfer learning algorithm is implemented. Using transfer learning, the knowledge learned from old data can be effectively used to solve the current prediction task, bridging the severe challenge brought about by long-term knowledge transfer. The distribution of time series data usually changes gradually and significantly over time; therefore, single-source transfer learning algorithm may also be confronted with the challenge of negative transfer. To solve this problem, *Gu & Dai (2021)* propose a new multi-source transfer learning algorithm, referred to as MultiSrcTL algorithm, and a new active multi-source transfer learning algorithm, referred to as AcMultiSrcTL algorithm.

Ye & Dai (2021) propose a deep transfer learning method (DTr-CNN) based on the CNN architecture, which inherits the advantages of CNN and tries to alleviate the problem of insufficient labeled data. This algorithm considers the similarity between the potential source dataset and the target dataset, and provides guidance for selecting the appropriate

source domain. *Gupta et al. (2018)* propose an approach to leverage deep RNNs for small, labeled datasets *via* transfer learning.

At present, there are relatively few studies on the application of transfer learning to time series prediction. Existing research mainly focuses on the research of pattern classification. In many practical situations, the lack of labeled data may become an obstacle to time series prediction. Unlike traditional machine learning algorithms, transfer learning breaks the assumption that training data and test data must follow the same distribution. For relevant datasets with sufficiently labeled samples, the use of transfer learning framework has become a new trend, and the use of knowledge from relevant source datasets on target dataset effectively solves the problem of insufficient labeled data.

Challenges and future trends

This section discusses the method of time series prediction. Time series data essentially reflects the changing trend of some random variables over time. The core of the time series prediction problem is to identify trends from the data, and use it to estimate the future data and predict occurrences in the next period of time. There is not one best model for all actual data, only the most suitable model from a reasonable range of models can be chosen to provide better prediction. The establishment of new time series models is still a problem that scholars will continue to study in the future, giving direction for further research in the field of time series prediction.

CONCLUSION

Time series is an important data type and is generated in almost every application domain at an unprecedented speed and scale. The analysis of time series can help us understand the essence of various phenomena. We investigate current research regarding time series and find that there are few reviews for time series algorithms. In this article, we analyze the prevalent topics of time series and divide them into two categories: classification and prediction. Further, we extract the important technology development routes for time series related algorithms, and introduce every original method and its subsequent improvements. In addition, we compare the performance of different algorithms, analyze and conclude their advantages and disadvantages, as well as the problems and challenges they face.

Through our investigation, we find that the technological development has three areas: the traditional method, machine learning method, and deep learning method. In time series classification, the mainstream methods change from distance-based methods (1NN-DTW) into feature-based methods (shapelets), and finally they evolve into a mathematical optimization problem that not only improve the accuracy but also reduce the time complexity. In time series prediction, owing to the limitations of AR, MA, ARIMA, and other traditional methods that cannot cope with nonlinear problems well, neural network methods have become a popular topic, and it is expected to enhance the learning ability of models through fuzzy cognitive map and transfer learning. Despite the fact that the current research has obtained some achievements, we find some important problems during our investigation:

- For time series classification, the research on semi-supervised and unsupervised learning algorithms is insufficient. While unlabeled data is ubiquitous and available in large amounts in real life, labeling it is labor intensive and sometimes requires expert knowledge.
- For time series prediction, constructing targeted time series models to solve real-world problems is still an ongoing problem for future researchers.

In view of the current development status of time series research, we believe that there are still many possible development directions for time series analysis. For example, neural network is a very popular method for time series analysis. In most cases, its solution process is a black box, which lacks interpretability, so that the results cannot be intuitively understood, and clear and targeted optimization scheme cannot be obtained. Exploring the symbolic expression of time series with stronger interpretability is the possible development direction of time series in the future. At present, most of the time series analysis is to collect data offline for offline analysis. When the model built in the offline phase is used in the online phase, new samples are continuously obtained as the working time increases. However, most methods do not consider the use of newly obtained data, and the model cannot be updated in time. Therefore, how to update the model for real-time data is the future task of time series modeling research.

Time series has attracted much attention because of its important applications in many fields, such as disease diagnosis and traffic flow prediction. We believe that the study of time series in this article will provide a valuable reference for related research and inspire interested researchers and practitioners to invest more in this promising field.

Funding
This research is supported by the Key-Area Research and Development Program of Guangdong Province (No. 2019B010107001), the fund of the National Engineering Laboratory for Agri-product Quality Traceability, Beijing Technology and Business University (No. AQT-2020-YB10) and the fund of Guangdong Support Program (No. 2019TY05X071). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures
The following grant information was disclosed by the authors:
The Key-Area Research and Development Program of Guangdong Province: No. 2019B010107001.
The National Engineering Laboratory for Agri-product Quality Traceability.
Beijing Technology and Business University: No. AQT-2020-YB10.
Guangdong Support Program: No. 2019TY05X071.

Competing Interests
Weijun Li is the co-founder of Shenzhen DAPU Microelectronics Co., Ltd.

Author Contributions

- Yuerong Tong performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Jingyi Liu performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Lina Yu conceived and designed the experiments, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Liping Zhang analyzed the data, prepared figures and/or tables, and approved the final draft.
- Linjun Sun analyzed the data, prepared figures and/or tables, and approved the final draft.
- Weijun Li conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Xin Ning performed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Jian Xu analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Hong Qin performed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Qiang Cai conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

Our paper does not contain raw data or code.

REFERENCES

Abanda A, Mori U, Lozano JA. 2019. A review on distance based time series classification. *Data Mining and Knowledge Discovery* **33(2)**:378–412 DOI 10.1007/s10618-018-0596-4.

Aghabozorgi S, Seyed Shirkhorshidi A, Ying Wah T. 2015. Time-series clustering – a decade review. *Information Systems* **53**:16–38 DOI 10.1016/j.is.2015.04.007.

Ali M, Alqahtani A, Jones MW, Xie X. 2019. Clustering and classification for time series data in visual analytics: a survey. *IEEE Access* **7**:181314–181338 DOI 10.1109/ACCESS.2019.2958551.

Athanasopoulos G, Vahid F. 2008. VARMA versus VAR for macroeconomic forecasting. *Journal of Business & Economic Statistics* **26(2)**:237–252 DOI 10.1198/073500107000000313.

Babu CN, Reddy BE. 2014. A moving-average filter based hybrid ARIMAANN model for forecasting time series data. *Applied Soft Computing* **23**:27–38 DOI 10.1016/j.asoc.2014.05.028.

Bagnall A, Lines J, Hills J, Bostrom A. 2015. Time-series classification with COTE: the collective of transformation-based ensembles. *IEEE Transactions on Knowledge and Data Engineering* **27(9)**:2522–2535 DOI 10.1109/TKDE.2015.2416723.

Baldán FJ, Bentez JM. 2018. Distributed fastshapelet transform: a big data time series classification algorithm. *Information Sciences* **496**:451–463 DOI 10.1016/j.ins.2018.10.028.

Bar-Joseph Z. 2004. Analyzing time series gene expression data. *Bioinformatics* **20(16)**:2493 DOI 10.1093/bioinformatics/bth283.

Barrera-Animas AY, Oyedele LO, Bilal M, Akinosho TD, Delgado JMD, Akanbi LA. 2022. Rainfall prediction: a comparative analysis of modern machine learning algorithms for time-series forecasting. *Machine Learning with Applications* **7**:100204 DOI 10.1016/j.mlwa.2021.100204.

Bates NR, Astor YM, Church MJ, Currie K, Dore JE, González-Dávila M, Lorenzoni L, Muller-Karger F, Olafsson J, Santana-Casiano JM. 2014. A time-series view of changing surface ocean chemistry due to ocean uptake of anthropogenic CO_2 and ocean acidification. *Oceanography* **27(1)**:126–141 DOI 10.2307/24862128.

Begum N, Hu B, Rakthanmanon T, Keogh E. 2013. Towards a minimum description length based stopping criterion for semi-supervised time series classification. In: *2013 IEEE 14th international conference on information reuse integration (IRI)*. 333–340 DOI 10.1109/IRI.2013.6642490.

Bose M, Mali K. 2019. Designing fuzzy time series forecasting models: a survey. *International Journal of Approximate Reasoning* **111**:78–99 DOI 10.1016/j.ijar.2019.05.002.

Box GEP, Jenkins G. 1970. Time series analysis forecasting and control. *Journal of Time Series Analysis* **3(3)**:131–133.

Chatfield C. 1978. The holt-winters forecasting procedure. *Journal of the Royal Statal Society* **27(3)**:264–279 DOI 10.2307/2347162.

Chen H-Y, Liu C-H, Sun B. 2017. Survey on similarity measurement of time series data mining. **32**:1–11 DOI 10.13195/j.kzyjc.2016.0462.

Chen T-L, Cheng C-H, Jong Teoh H. 2007. Fuzzy time-series based on Fibonacci sequence for stock price forecasting. *Physica A: Statistical Mechanics and its Applications* **380**:377–390 DOI 10.1016/j.physa.2007.02.084.

Chen Y, Hu B, Keogh E, Batista GE. 2013. DTW-D: time series semi-supervised learning from a single example. In: *Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining. KDD '13*. New York, NY, USA: Association for Computing Machinery, 383–391 DOI 10.1145/2487575.2487633.

Chung Fu T. 2011. A review on time series data mining. *Engineering Applications of Artificial Intelligence* **24(1)**:164–181 DOI 10.1016/j.engappai.2010.09.007.

Connor J, Atlas L. 1991. Recurrent neural networks and time series prediction. In: *Ijcnn-91-Seattle international joint conference on neural networks*. Piscataway: IEEE 301–306 DOI 10.1109/IJCNN.1991.155194.

Corradi V. 1995. Nonlinear transformations of integrated time series: a reconsideration. *Journal of Time Series Analysis* **16(6)**:539–549 DOI 10.1111/j.1467-9892.1995.tb00253.x.

Dau HA, Silva DF, Petitjean F, Forestier G, Keogh E. 2018. Optimizing dynamic time warping's window width for time series data mining applications. *Data Mining and Knowledge Discovery* **32(5)**:1–47 DOI 10.1007/s10618-017-0516-z.

De Gooijer JG, Hyndman RJ. 2006. 25 years of time series forecasting. *International Journal of Forecasting* **22**:443–473.

de O. Santos Jnior DS, de Oliveira JF, de Mattos Neto PS. 2019. An intelligent hybridization of ARIMA with machine learning models for time series forecasting. *Knowledge-Based Systems* **175**:72–86 DOI 10.1016/j.knosys.2019.03.011.

Deb C, Zhang F, Yang J, Lee SE, Shah KW. 2017. A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews* **74**:902–924 DOI 10.1016/j.rser.2017.02.085.

Dempster A, Petitjean F, Webb GI. 2019. ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels. *ArXiv preprint* DOI 10.48550/arXiv.1910.13051.

Ding H, Trajcevski G, Scheuermann P, Wang X, Keogh E. 2008. Querying and mining of time series data: experimental comparison of representations and distance measures. *Proceedings of the VLDB Endowment* **1(2)**:15421552 DOI 10.14778/1454159.1454226.

Fakhrazari A, Vakilzadian H. 2017. A survey on time series data mining. In: *IEEE International conference on electro information technology (EIT)*. IEEE, 476–481 DOI 10.1109/EIT.2017.8053409.

Fawaz HI, Forestier G, Weber J, Idoumghar L, Muller PA. 2019. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery* **33(4)**:917–963 DOI 10.1007/s10618-019-00619-1.

Felix G, Npoles G, Falcon R, Froelich W, Bello R. 2019. A review on methods and software for fuzzy cognitive maps. *Artificial Intelligence Review* **52(3)**:1707–1737 DOI 10.1007/s10462-017-9575-1.

Froelich W, Papageorgiou EI, Samarinas M, Skriapas K. 2012. Application of evolutionary fuzzy cognitive maps to the long-term prediction of prostate cancer. *Applied Soft Computing* **12(12)**:3810–3817 DOI 10.1016/j.asoc.2012.02.005.

Gao R, Du L, Yuen KF. 2020. Robust empirical wavelet fuzzy cognitive map for time series forecasting. *Engineering Applications of Artificial Intelligence* **96**:103978 DOI 10.1016/j.engappai.2020.103978.

Gardner ES. 1985. Exponential smoothing: the state of the art. *Journal of Forecasting* **4(1)**:1–28 DOI 10.1002/for.3980040103.

Geurts P. 2001. Pattern extraction for time series classification. In: *Proceedings of the 5th European conference on principles of data mining and knowledge discovery. PKDD '01*. Berlin, Heidelberg: Springer-Verlag, 115–127 DOI 10.1007/3-540-44794-6_10.

Ghalwash MF, Obradovic Z. 2012. Early classification of multivariate temporal observations by extraction of interpretable shapelets. *BMC Bioinformatics* **13(1)**:195 DOI 10.1186/1471-2105-13-195.

Ghil M, Allen MR, Dettinger MD, Ide K, Kondrashov D, Mann ME, Robertson AW, Saunders A, Tian Y, Varadi Fa. 2002. Advanced spectral methods for climatic time series. *Reviews of Geophysics* **40(1)**:3-1–3-41 DOI 10.1029/2000RG000092.

Gluhovsky A, Agee E. 2007. On the analysis of atmospheric and climatic time series. *Journal of Applied Meteorology & Climatology* **46(7)**:1125–1129 DOI 10.1175/JAM2512.1.

Göb R. 2006. Data mining and statistical control - a review and some links. *Physica-Verlag HD* **8**:285–308 DOI 10.1007/3-7908-1687-6_17.

Grabocka J, Schilling N, Wistuba M, Schmidt-Thieme L. 2014. Learning time-series shapelets. In: *Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining. KDD '14.* New York, NY, USA: Association for Computing Machinery, 392–401 DOI 10.1145/2623330.2623613.

Gu Q, Dai Q. 2021. A novel active multi-source transfer learning algorithm for time series forecasting. *Applied Intelligence* **51(3)**:1326–1350 DOI 10.1007/s10489-020-01871-5.

Gupta P, Malhotra P, Vig L, Shroff G. 2018. Transfer Learning for Clinical Time Series Analysis using Recurrent Neural Networks. *ArXiv preprint* DOI 10.48550/arXiv.1807.01705.

Hajirahimi Z, Khashei M. 2019. Hybrid structures in time series modeling and forecasting: a review. *Engineering Applications of Artificial Intelligence* **86**:83–106 DOI 10.1016/j.engappai.2019.08.018.

He G, Duan Y, Peng R, Jing X, Qian T, Wang L. 2015. Early classification on multivariate time series. *Neurocomput* **149(PB)**:777–787 DOI 10.1016/j.neucom.2014.07.056.

He G, Zhao W, Xia X. 2020. Confidence-based early classification of multivariate time series with multiple interpretable rules. *Pattern Analysis and Applications* **23(2)**:567–580 DOI 10.1007/s10044-019-00782-7.

He G, Zhao W, Xia X, Peng R, Wu X. 2019. An ensemble of shapelet-based classifiers on inter-class and intra-class imbalanced multivariate time series at the early stage. *Soft Comput.* **23(15)**:6097–6114 DOI 10.1007/s00500-018-3261-3.

Hills J, Lines J, Baranauskas E, Mapp J, Bagnall A. 2014. Classification of time series by shapelet transformation. *Data Mining and Knowledge Discovery* **28(4)**:851–881 DOI 10.1007/s10618-013-0322-1.

Hochreiter S, Schmidhuber J. 1996. Bridging long time lags by weight guessing and "long short term memory". *Spatiotemporal Models in Biological and Artificial Systems* **37(65-72)**:11.

Holt CC. 2004. Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting* **20(1)**:5–10 DOI 10.1016/j.ijforecast.2003.09.015.

Homenda W, Jastrzebska A, Pedrycz W. 2015. Joining concept's based fuzzy cognitive map model with moving window technique for time series modeling. In: *IFIP*

international conference on computer information systems and industrial management. Berlin, Heidelberg: Springer, 397–408.

Hou L, Kwok JT, Zurada JM. 2016. Efficient learning of timeseries shapelets. In: *Proceedings of the thirtieth AAAI conference on artificial intelligence. AAAI'16.* Cambridge: AAAI Press, 1209–1215.

Idrees SM, Alam MA, Agarwal P. 2019. A prediction approach for stock market volatility based on time series data. *IEEE Access* **7**:17287–17298 DOI 10.1109/ACCESS.2019.2895252.

Iwana BK, Frinken V, Uchida S. 2020. DTW-NN: a novel neural network for time series recognition using dynamic alignment between inputs and weights. *Knowledge-Based Systems* **188**:104971 DOI 10.1016/j.knosys.2019.104971.

Jiang QY. 2015. Time series prediction based on machine learning. In: *2015 International conference on electrical, automation and mechanical engineering.* Atlantis Press, 128–129.

Kadous MW, Sammut C. 2005. Classification of multivariate time series and structured data using constructive induction. *Machine Learning* **58(2/3)**:179–216 DOI 10.1007/s10994-005-5826-5.

Kalpakis K, Gada D, Puttagunta V. 2001. Distance measures for effective clustering of ARIMA time-series. In: *Proceedings 2001 IEEE international conference on data mining.* Piscataway: IEEE, 273–280 DOI 10.1109/ICDM.2001.989529.

Khashei M, Bijari M. 2010. An artificial neural network (p,d,q) model for timeseries forecasting. *Expert Systems with Applications* **37(1)**:479–489 DOI 10.1016/j.eswa.2009.05.044.

Kim YJ. 2017. Comparison between inverse model and chaos time series inverse model for long-term prediction. *Sustainability* **9(6)**:982 DOI 10.3390/su9060982.

Kosko B. 1986. Fuzzy cognitive maps. *International Journal of Man-Machine Studies* **24(1)**:65–75 DOI 10.1016/S0020-7373(86)80040-2.

Langkvist M, Karlsson L, Loutfi A. 2014. A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters* **42**:11–24 DOI 10.1016/j.patrec.2014.01.008.

Lara-Benítez P, Carranza-García M, Riquelme JC. 2021. An experimental review on deep learning architectures for time series forecasting. *International Journal of Neural Systems* **31(03)**:2130001 DOI 10.1142/S0129065721300011.

Li D, Wang Y, Xu B, Li W, Li W, Yu L, Yang Q. 2019. PIPULS: predicting I/O patterns using LSTM in storage systems. In: *2019 international conference on high performance big data and intelligent systems (HPBD IS).* 14–21 DOI 10.1109/HPBDIS.2019.8735467.

Li X, Wu P, Wang W. 2020. Incorporating stock prices and news sentiments for stock market prediction: a case of Hong Kong. *Information Processing & Management* **57(5)**:102212 DOI 10.1016/j.ipm.2020.102212.

Lim B, Zohren S. 2021. Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society a Mathematical Physical and Engineering Sciences* **379(2194)**:20200209 DOI 10.1098/rsta.2020.0209.

Lin YF, Chen HH, Tseng VS, Pei J. 2015. Reliable early classification on multivariate time series with numerical and categorical attributes. In: *Pacific-Asia conference on knowledge discovery and data mining*. Cham: Springer, 199–211.

Lines J, Davis LM, Hills J, Bagnall A. 2012. A Shapelet transform for time series classification. In: *Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining. KDD '12*. New York, NY, USA: Association for Computing Machinery, 289–297 DOI 10.1145/2339530.2339579.

Liu H, Yan G, Duan Z, Chen C. 2021. Intelligent modeling strategies for forecasting air quality time series: a review. *Applied Soft Computing* **102**:106957 DOI 10.1016/j.asoc.2020.106957.

Liu J, Zhao K, Kusy B, Wen J, Jurdak R. 2015. Temporal embedding in convolutional neural networks for robust learning of abstract snippets. *ArXiv preprint* DOI 10.48550/arXiv.1502.05113.

Lopez BJ, Steven C, Antonio G. 2017. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *International Journal of Epidemiology* **46(1)**:348–355 DOI 10.1093/ije/dyw098.

Lu W, Yang J, Liu X. 2014. The hybrids algorithm based on fuzzy cognitive map for fuzzy time series prediction. *Journal of Information & Computational Ence* **11(2)**:357–366 DOI 10.12733/jics20102682.

Lu W, Yang J, Liu X, Pedrycz W. 2014. The modeling and prediction of time series based on synergy of high-order fuzzy cognitive map and fuzzy c-means clustering. *Knowledge Based Systems* **70(nov.)**:242–255 DOI 10.1016/j.knosys.2014.07.004.

Mahalakshmi G, Sridevi S, Rajaram S. 2016. A survey on forecasting of time series data. In: *2016 international conference on computing technologies and intelligent data engineering (ICCTIDE'16)*. 1–8 DOI 10.1109/ICCTIDE.2016.7725358.

Marussy K, Buza K. 2013. Success: a new approach for semi-supervised classification of time-series. In: *International conference on artificial intelligence and soft computing*. Berlin, Heidelberg: Springer, 437–447 DOI 10.1007/978-3-642-38658-9_39.

Maçaira PM, Tavares Thom AM, Cyrino Oliveira FL, Carvalho Ferrer AL. 2018. Time series analysis with explanatory variables: a systematic literature review. *Environmental Modelling & Software* **107**:199–209 DOI 10.1016/j.envsoft.2018.06.004.

McDonald , Alan J. 1986. Periodic smoothing of time series. *Siam Journal on Scientific & Statistical Computing* **7(2)**:665–688 DOI 10.1137/0907045.

Mehrmolaei S, Keyvanpourr MR. 2015. A brief survey on event prediction methods in time series. In: Silhavy R, Senkerik R, Oplatkova ZK, Prokopova Z, Silhavy P, eds. *Artificial intelligence perspectives and applications.* Cham: Springer International Publishing, 235–246 DOI 10.1007/978-3-319-18476-0_24.

Misra P, Siddharth . 2017. Machine learning and time series: real world applications. In: Astya P, Swaroop A, Sharma V, Singh M, Gupta K, eds. *2017 IEEE international conference on computing, communication and automation (ICCCA)*. Piscataway: IEEE, 389–394 DOI 10.1109/CCAA.2017.8229832.

Mittelman R. 2015. Time-series modeling with undecimated fully convolutional neural networks. *ArXiv preprint* DOI 10.48550/ARXIV.1508.00317.

Mizon CGE. 1991. Empirical analysis of macroeconomic time series: VAR and structural models. *European Economic Review* **35(4)**:887–917 DOI 10.1016/0014-2921(91)90042-H.

Mori U, Mendiburu A, Dasgupta S, Lozano JA. 2018. Early classification of time series by simultaneously optimizing the accuracy and earliness. *IEEE Transactions on Neural Networks and Learning Systems* **29(10)**:4569–4578 DOI 10.1109/TNNLS.2017.2764939.

Mori U, Mendiburu A, Keogh E, Lozano JA. 2016. Reliable early classification of time series based on discriminating the classes over time. *Data Mining & Knowledge Discovery* **31(1)**:1–31 DOI 10.1007/s10618-016-0462-1.

Mueen A, Keogh E, Young N. 2011. Logical-shapelets: an expressive primitive for time series classification. In: *Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining. KDD '11.* New York, NY, USA: Association for Computing Machinery, 1154–1162 DOI 10.1145/2020408.2020587.

Namoano B, Starr A, Emmanouilidis C, Cristobal RC. 2019. Online change detection techniques in time series: an overview. In: *2019 IEEE international conference on prognostics and health management (ICPHM).* Piscataway: IEEE, 1–10 DOI 10.1109/ICPHM.2019.8819394.

Nguyen MN, Li X-L, Ng S-K. 2011. Positive unlabeled learning for time series classification. In: *Proceedings of the twenty-second international joint conference on artificial intelligence - volume volume two, IJCAI'11.* AAAI Press, 1421–1426.

Nordman DJ, Lahiri SN. 2014. A review of empirical likelihood methods for time series. *Journal of Statistical Planning and Inference* **155**:1–18 DOI 10.1016/j.jspi.2013.10.001.

Nyein Naing WY, Htikemuhammad Yusof ZZ. 2015. State of the art machine learning techniques for time series forecasting: a survey. *Advanced Ence Letters* **21(11)**:3574–3576 DOI 10.1166/asl.2015.6586.

Oliveira JaFLd, Ludermir TB. 2014. A hybrid evolutionary system for parameter optimization and lag selection in time series forecasting. In: *Proceedings of the 2014 Brazilian conference on intelligent system. BRACIS '14.* USA: IEEE Computer Society, 73–78 DOI 10.1109/BRACIS.2014.24.

Osmanoglu B, Sunar F, Wdowinski S, Cabral-Cano E. 2016. Time series analysis of InSAR data: methods and trends. *Isprs Journal of Photogrammetry & Remote Sensing* **115(may)**:90–102 DOI 10.1016/j.isprsjprs.2015.10.003.

Pai P-F, Lin C-S. 2005. A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega* **33(6)**:497–505 DOI 10.1016/j.omega.2004.07.024.

Pan SJ, Yang Q. 2010. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering* **22(10)**:1345–1359 DOI 10.1109/TKDE.2009.191.

Papageorgiou EI, Froelich W. 2012a. Application of evolutionary fuzzy cognitive maps for prediction of pulmonary infections. *IEEE Transactions on Information Technology in Biomedicine* **16(1)**:143–149 DOI 10.1109/TITB.2011.2175937.

Papageorgiou EI, Froelich W. 2012b. Multi-step prediction of pulmonary infection with the use of evolutionary fuzzy cognitive maps. *Neurocomputing* **92**:28–35 DOI 10.1016/j.neucom.2011.08.034.

Papageorgiou EI, Poczeta K, Laspidou C. 2015. Application of fuzzy cognitive maps to water demand prediction. In: *2015 IEEE international conference on fuzzy systems (FUZZ-IEEE)*. Piscataway: IEEE, 1–8 DOI 10.1109/FUZZ-IEEE.2015.7337973.

Papo D. 2021. Algorithmic approaches for assessing irreversibility in time series: review and comparison. *Entropy* **23(11)**:1474 DOI 10.3390/e23111474.

Patton AJ. 2012. A review of copula models for economic time series. *Journal of Multivariate Analysis* **110**:418 DOI 10.1016/j.jmva.2012.02.021.

Peter G, Zhang . 2003. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* **50**:159–175 DOI 10.1016/S0925-2312(01)00702-0.

Poczeta K, Yastrebov A. 2014. Analysis of fuzzy cognitive maps with multi-step learning algorithms in valuation of owner-occupied homes. In: *2014 IEEE international conference on fuzzy systems (FUZZ-IEEE)*. Piscataway: IEEE, 1029–1035 DOI 10.1109/FUZZ-IEEE.2014.6891587.

Poczeta K, Yastrebov A, Papageorgiou EI. 2015. Learning fuzzy cognitive maps using structure optimization genetic algorithm. In: *2015 federated conference on computer science and information systems (FedCSIS)*. IEEE, 547–554 DOI 10.15439/2015F296.

Protheroe RJ, Hocking WK. 1988. Period searches using the fast Fourier transform in X-ray and gamma-ray astronomy. *Astrophysical Letters & Communications* **27**:237–240.

Radha P, Divya R. 2017. Survey on the multiple time series data with data mining techniques. *International Journal of Computational Intelligence Research* **13(4)**:615–620.

Rakthanmanon T, Campana B, Mueen A, Batista G, Westover B, Zhu Q, Zakaria J, Keogh E. 2012. Searching and mining trillions of time series subsequences under dynamic time warping. In: *Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining. KDD '12*. New York, NY, USA: Association for Computing Machinery, 262–270 DOI 10.1145/2339530.2339576.

Rakthanmanon T, Keogh E. 2013. Fast shapelets: a scalable algorithm for discovering time series shapelets. In: *Proceedings of the 2013 SIAM international conference on data mining*. 668–676 DOI 10.1137/1.9781611972832.74.

Rani S, Sikka G. 2012. Recent techniques of clustering of time series data: a survey. *International Journal of Computer Applications* **52(15)**:1–9.

Rodriguez JJ, Alonso CJ, Bostrom H. 2001. Boosting interval based literals. *Intelligent Data Analysis* **5(3)**:245–262(18) DOI 10.3233/IDA-2001-5305.

Salles R, Belloze K, Porto F, Gonzalez PH, Ogasawara E. 2019. Nonstationary time series transformation methods: an experimental review. *Knowledge-Based Systems* **164**:274–291 DOI 10.1016/j.knosys.2018.10.041.

Salmeron JL, Mansouri T, Moghadam MRS, Mardani A. 2019. Learning fuzzy cognitive maps with modified asexual reproduction optimisation algorithm. *Knowledge Based Systems* **163(JAN.1)**:723–735 DOI 10.1016/j.knosys.2018.09.034.

Salmeron JL, Rahimi SA, Navali AM, Sadeghpour A. 2017. Medical diagnosis of rheumatoid arthritis using data driven PSO-FCM with scarce datasets. *Neurocomputing* **232**:104–112 DOI 10.1016/j.neucom.2016.09.113.

Salmeron JL, Ruiz-Celma A, Mena A. 2016. Learning FCMs with multi-local and balanced memetic algorithms for forecasting industrial drying processes. *Neurocomputing* **232**:52–57 DOI 10.1016/j.neucom.2016.10.070.

Sang Y-F. 2013. A review on the applications of wavelet transform in hydrology time series analysis. *Atmospheric Research* **122**:8–15 DOI 10.1016/j.atmosres.2012.11.003.

Sart D, Mueen A, Najjar W, Keogh E, Niennattrakul V. 2010. Accelerating dynamic time warping subsequence search with GPUs and FPGAs. In: *Proceedings of the 2010 IEEE international conference on data mining. ICDM '10.* USA: IEEE Computer Society, 1001–1006 DOI 10.1109/ICDM.2010.21.

Schreiber T. 1999. Interdisciplinary application of nonlinear time series methods. *Physics Reports* **308(1)**:1–64 DOI 10.1016/S0370-1573(98)00035-0.

Scotto MG, Wei CH, Gouveia S. 2015. Thinning-based models in the analysis of integer-valued time series: a review. *Statistical Modelling* **15(6)**:590–618 DOI 10.1177/1471082X15584701.

Seyedjamal Z, Saeed A, Wah TY. 2014. A review of subsequence time series clustering. *The Scientific World Journal* **2014**:312521 DOI 10.1155/2014/312521.

Sezer OB, Gudelek MU, Ozbayoglu AM. 2020. Financial time series forecasting with deep learning : a systematic literature review: 2005–2019. *Applied Soft Computing* **90**:106181 DOI 10.1016/j.asoc.2020.106181.

Shahane AN, Thomas D, Bock P. 1977. Spectral analysis of hydrometeorological time series. *Water Resources Research* **13(1)**:41–49 DOI 10.1029/WR013i001p00041.

Shen Z, Zhang Y, Lu J, Xu J, Xiao G. 2020. A novel time series forecasting model with deep learning. *Neurocomputing* **396**:302–313 DOI 10.1016/j.neucom.2018.12.084.

Slutzky E. 1937. The summation of random causes as the source of cyclic processes. *Econometrica* **5(2)**:105–146 DOI 10.2307/1907241.

Song P, Russell CT. 1999. Time series data analyses in space physics. *Space Science Reviews* **87(3–4)**:387–463 DOI 10.1023/A:1005035800454.

Spiro AG, D GM, Kiseleva NE, Pokrovskaya IV. 2018. Segmentation and hashing of time series in stock market prediction. *Automation & Remote Control* **79(5)**:911–918 DOI 10.1134/S0005117918050119.

Stach W, Kurgan L, Pedrycz W, Reformat M. 2005. Genetic learning of fuzzy cognitive maps. *Fuzzy Sets & Systems* **153(3)**:371–401 DOI 10.1016/j.fss.2005.01.009.

Suykens , Johan AK. 2002. Least squares support vector machines. *International Journal of Circuit Theory & Applications* **27(6)**:605–615.

Talaei-Khoei A, Wilson JM. 2019. Using time-series analysis to predict disease counts with structural trend changes. *Information Processing & Management* **56(3)**:674–686 DOI 10.1016/j.ipm.2018.11.004.

Tang L, Lv H, Yang F, Yu L. 2015. Complexity testing techniques for time series data: A comprehensive literature review. *Chaos, Solitons & Fractals* **81**:117–135 DOI 10.1016/j.chaos.2015.09.002.

Tealab A. 2018. Time series forecasting using artificial neural networks methodologies: a systematic review. *Future Computing and Informatics Journal* **3(2)**:334–340 DOI 10.1016/j.fcij.2018.10.003.

Teichgraeber H, Brandt AR. 2022. Time-series aggregation for the optimization of energy systems: goals, challenges, approaches, and opportunities. *Renewable and Sustainable Energy Reviews* **157**:111984 DOI 10.1016/j.rser.2021.111984.

Vanhoenshoven F, Npoles G, Froelich W, Salmeron JL, Vanhoof K. 2020. Pseudoinverse learning of fuzzy cognitive maps for multivariate time series forecasting. *Applied Soft Computing* **95**:106461 DOI 10.1016/j.asoc.2020.106461.

Wang H, Zhang Q, Wu J, Pan S, Chen Y. 2019. Time series feature learning with labeled and unlabeled data. *Pattern Recognition* **89**:55–66 DOI 10.1016/j.patcog.2018.12.026.

Wang L, Zeng Y, Chen T. 2015. Back propagation neural network with adaptive differential evolution algorithm for time series forecasting. *Expert Systems with Applications* **42(2)**:855–863 DOI 10.1016/j.eswa.2014.08.018.

Wangdi K, Singhasivanon P, Silawan T, Lawpoolsri S, White NJ, Kaewkungwal J. 2010. Development of temporal modelling for forecasting and prediction of malaria infections using time-series and ARIMAX analyses: a case study in endemic districts of Bhutan. *Malaria Journal* **9(1)**:251 DOI 10.1186/1475-2875-9-251.

Wei L, Keogh E. 2006. Semi-supervised time series classification. In: *Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining. KDD '06*. New York, NY, USA: Association for Computing Machinery, 748–753 DOI 10.1145/1150402.1150498.

Wilson D, Passmore S, Tang Y, Vanzwieten J. 2018. Bidirectional long short-term memory networks for rapid fault detection in marine hydrokinetic turbines. In: *2018 17th IEEE international conference on machine learning and applications (ICMLA)*. Piscataway: IEEE, 495–500 DOI 10.1109/ICMLA.2018.00080.

Winters PR. 1960. Forecasting sales by exponentially weighted moving averages. *Management Science* **6(3)**:324–342 DOI 10.1287/mnsc.6.3.324.

Wistuba M, Grabocka J, Schmidt-Thieme L. 2015. Ultra-fast shapelets for time series classification. *ArXiv preprint* DOI 10.48550/arXiv.1503.05018.

Wu R, Der A, Keogh EJ. 2021. When is early classification of time series meaningful? *IEEE Transactions on Knowledge and Data Engineering*. 1–1 DOI 10.1109/TKDE.2021.3108580.

Xi X, Keogh E, Shelton C, Wei L, Ratanamahatana CA. 2006. Fast time series classification using numerosity reduction. In: *Proceedings of the 23rd international conference on Machine learning*. 1033–1040 DOI 10.1145/143844.1143974.

Xiao J, He C, Wang S. 2012. Crude oil price forecasting: a transfer learning based analog complexing model. In: *2012 Fifth international conference on business intelligence and financial engineering*. IEEE, 29–33 DOI 10.1109/BIFE.2012.14.

Xing Z, Pei J, Yu PS. 2009. Early prediction on time series: a nearest neighbor approach. In: *Proceedings of the 21st international jont conference on artifical intelligence. IJCAI'09*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1297–1302.

Xing Z, Pei J, Yu PS. 2012. Early classification on time series. *Knowledge & Information Systems* **31(1)**:105–127 DOI 10.1007/s10115-011-0400-x.

Xing Z, Pei J, Yu PS, Wang K. 2011. Extracting interpretable features for early classification on time series. In: *Proceedings of the eleventh SIAM international conference on data mining.* Society for Industrial and Applied Mathematics, 247–258 DOI 10.1137/1.9781611972818.22.

Xu-Dong S. 2019. A survey of time series algorithms based on deep learning. *Information Technology and Informatization* **1**:71–76.

Xu Z, Funaya K. 2015. Time series analysis with graph-based semi-supervised learning. In: *2015 IEEE international conference on data science and advanced analytics (DSAA).* Piscataway: IEEE, 1–6 DOI 10.1109/DSAA.2015.7344902.

Yang S, Liu J. 2018. Time series forecasting based on high-order fuzzy cognitive maps and wavelet transform. *IEEE Transactions on Fuzzy Systems* **26(6)**:3391–3402 DOI 10.1109/TFUZZ.2018.2831640.

Yang Q, Wu X. 2006. 10 challenging problems in data mining research. *International Journal of Information Technology & Decision Making* **5(04)**:597–604 DOI 10.1142/S0219622006002258.

Ye R, Dai Q. 2018. A novel transfer learning framework for time series forecasting. *Knowledge-Based Systems* **156**:74–99 DOI 10.1016/j.knosys.2018.05.021.

Ye R, Dai Q. 2021. Implementing transfer learning across different datasets for time series forecasting. *Pattern Recognition* **109**:107617 DOI 10.1016/j.patcog.2020.107617.

Ye L, Keogh E. 2009. Time series shapelets: A new primitive for data mining. In: *Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining. KDD '09.* New York, NY, USA: Association for Computing Machinery, 947–956 DOI 10.1145/1557019.1557122.

Ye J, Zhao J, Ye K, Xu C. 2020. Multi-STGCnet: a graph convolution based spatial-temporal framework for subway passenger flow forecasting. In: *2020 International joint conference on neural networks (IJCNN).* 1–8 DOI 10.1109/IJCNN48605.2020.9207049.

Yesil E, Ozturk C, Dodurka MF, Sakalli A. 2013. Fuzzy cognitive maps learning using Artificial Bee Colony optimization. In: *2013 IEEE international conference on fuzzy systems (FUZZ-IEEE).* Piscataway: IEEE, 1–8 DOI 10.1109/FUZZ-IEEE.2013.6622524.

Yule GU. 1927. On a method of investigating periodicities in disturbed series, with special reference to wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society B Biological Sciences* **226(636–646)**:267–298.

Zeger SL, Irizarry R, Peng RD. 2006. On time series analysis of public health and biomedical data. *Annual Review of Public Health* **27(1)**:57 DOI 10.1146/annurev.publhealth.26.021304.144517.

Zhao H, Pan Z, Tao W. 2020. Regularized shapelet learning for scalable time series classification. *Computer Networks* **173**:107171 DOI 10.1016/j.comnet.2020.107171.

Zheng Y, Liu Q, Chen E, Ge Y, Zhao JL. 2014. Time series classification using multi-channels deep convolutional neural networks. In: *Web-age information management.* Cham: Springer International Publishing, 298–310 DOI 10.1007/978-3-319-08010-9_33.

Zheng-Xin LI, Feng-Ming Z, Xiao-Feng Z, Ji-Cheng C, Chao LI. 2017. Survey of similarity search for multivariate time series. *Control and Decision* **32(4):**577–583.

Zhu Z. 2017. Change detection using landsat time series: a review of frequencies, preprocessing, algorithms, and applications. *Isprs Journal of Photogrammetry & Remote Sensing* **130(aug.):**370–384 DOI 10.1016/j.isprsjprs.2017.06.013.