WILEY | Hindawi

Review Article

Time Series Based Data Explorer and Stream Analysis for Anomaly Prediction

Xiao-Xia Yin ⓘ,[1,2] Yuan Miao,[2] and Yanchun Zhang ⓘ[1,3]

[1]*Cyberspace Institute of Advanced Technology, Guangzhou University, Guangzhou 510006, China*
[2]*College of Engineering and Science, Victoria University, VIC 8001, Australia*
[3]*Pengcheng Laboratory, Shenzhen, China*

Correspondence should be addressed to Yanchun Zhang; yanchun.zhang@gzhu.edu.cn

All over the world, time series-based anomaly prediction plays a vital role in all walks of life such as medical monitoring in hospitals and climate and environment risks. In the present study, a survey on the methods and techniques for time series data mining and proposes is carried, in order to solve a brand-new problem, time series progressive anomaly prediction. In terms of contents, the first part sketches out the methods that have captured most of the interest of researchers, which include an overview of abnormal prediction problems, a summary of main characteristics of anomaly prediction, and an introduction of anomaly prediction methodology in literature. The second part focuses on the future research trends on the phase/staged abnormal prediction of time series, where a novel time series compression method and a corresponding similarity measure will be designed, which can be explored subsequently. Finally, the related challenges to take this trend are mentioned. It is hoped that this paper can provide a profound understanding of anomaly prediction for the time series-based data mining research field.

1. Introduction

Time series, a type of data widely existing in production and life, is widely applied in the fields such as medical monitoring [1–3], environmental monitoring [4, 5], and stock tickers [6, 7]. In the intensive care environment, online automatic abnormality analysis of multidimensional temporal data streams assists medical staff in grasping the health risks encountered by patients, which can benefit patients and save their lives [1] and increase the survival rate of casualties in accidents [2]. The mining of long-term electrocardiograms (for several days) provides a new biomarker for predicting the mortality risk of patients with heart disease patients [3]. In environmental monitoring, water quality data streams can be used for pollution monitoring [4], and meteorological data stream can be used for extreme weather predictions [5].

Anomaly monitoring based on time series is a typical example of data stream applications, which can be divided into two types, namely, anomaly detection of current or his-

torical data and prediction of potential anomalies. There are practical limitations when only detecting current/historical anomalies, and such limitations can be broken through by the prediction of anomalies. For example, in the field of medical and health care, early detection of certain diseases is of great importance to improve therapeutic effects and eradicate these diseases [8]; but under current medical conditions, diseases are generally treated after symptoms are detected.

Furthermore, treatment is ended upon the disappearance of the symptoms that can be judged by online medical data streams, without concerning the possible rehospitalization. Studies have shown that approximately 20 percent of patients are readmitted within 30 days after the discharge, and 35 percent of patients are readmitted within 90 days after the discharge. This leads to high expenditures, such as US$17.4 billion per year [9], and better disease prediction can help save costs. For another example, empirical research [10] showed that approximately 0.9 percent of tornadoes were not predicted, and the proportion of deaths in these

tornadoes is up to 8.5 percent; however, another study [11] showed that, the early warning in 6–10 minutes before a tornado can comparatively reduce deaths by 41 percent.

As an important part of data mining, mining time series data streams has been widely explored by researchers [12–41]. Among the existing research results, a few studies [3, 5, 27–30, 42, 43] are related to the problem of abnormal prediction [3]. For instance, ECG mining is used to predict the mortality risk of patients [27], optimization methods are used to evaluate the aging degree of systems, and early time series classification is performed. However, the characteristics of anomaly prediction algorithms have not been discussed deeply.

Generally, anomaly detection using forecasting is viewed as a technique to generate a forecast of an unusual point or a single instance in a given set that is different from others owing to its attributes. If the data is anomalous in some context, it is called contextual anomaly.

In the current situation of COVID-19, the whole world is experiencing unprecedented scenarios everywhere, from medical supplies running short, healthcare systems getting overwhelmed, to fears of an economic downfall. Often, the unexpected outbreak is termed as the "new normal" of life with COVID-19. Before becoming the "new normal", it is necessary to improve the ability for the accurate prediction of future disease spread, and effective analysis of the death and recovery rates, so as to better understand the current situation, discover insights on the future development of the disease, and thus allowing humanity to make better preparations.

This survey aims to solve a brand-new problem existing in time series mining, namely, time series progressive anomaly prediction. With this regard, the various stages of the anomaly evolution process shall be identified automatically, and the type of anomaly and the time when the anomaly occurs at each stage shall be predicated. On this basis, we propose to design an effective time series progressive anomaly prediction algorithm. Each stage of the anomaly evolution route can be represented by a set of characteristic subsequences. A dynamic segmentation scheme will be utilized to segment training series [44, 45], which is iteratively combined with the identification of characteristic subsequences, thus yielding optimal representation for each latent stage on the evolution route. Then, a rule set will be constructed for online prediction. The research mainly involves the design of an effective time series compression method, a similarity measure upon the compressed sequences, and an effective time series progressive anomaly prediction method.

In the present study, we intend to analyze the existing research results and time sequence analysis methodology. Meanwhile, we clearly describe that the time sequence anomaly prediction should have some characteristics, and describe the overall anomaly prediction algorithms. The structure of this paper is as follows: Section 2 describes the anomaly prediction methodology in literature. Section 3 outlines the main mathematical background using time series mining methods to carry out anomaly prediction. Section 4 provides a general overview of the characteristics of anomaly prediction. Section 5 shows significant suggestions on the staged abnormal prediction of time series, which can be explored subsequently. Section 6 represents technical challenges in multiple data streams analysis for anomaly prediction. Section 7 is related to the design of compression and similarity measurement frameworks to achieve efficient phase anomaly prediction algorithm, thus laying a foundation for future research in time series stream analysis.

2. Anomaly Prediction Methodology in Literature

At present, most of the research on anomaly prediction [3, 27–30, 42, 43] focuses on solving early classification [5, 28–30, 42, 43] of time series. Namely, a complete (univariate or multivariate) time series is classified only by the time series prefix composed of the observation values of the first several time stamps. This method satisfies our requirements for early prediction of abnormalities (i.e., detecting abnormal signs as soon as possible) but cannot meet the requirements for the phase abnormal prediction (phase division and continuous monitoring of abnormal evolution paths); other related research results [3, 27] also have similar problems. Illustratively, [3] ECG mining is applied to predict the mortality risk of patients but does not continuously monitor and stage patients' disease evolution process; [27] solve the system problem of evaluating the degree of aging, which considers the stages of the aging process to a certain extent, but for each stage, it fails to estimate the time interval from the occurrence of the abnormality. How to effectively divide and characterize the evolution stage of anomalous signs is a problem demanding prompt solution. Without paying enough attention to the above problems, the existing research results still provide us with valuable solution references. Considering anomaly prediction is deemed as a classification problem, the early classification of time series has become research priority [5, 28–30, 42, 43]. The existing early classification research can be divided into two categories: one is the classification method based on global features [28, 42, 43], where each observation in the complete sequence (or sequence prefix) is regarded as a part of the feature; the other is classification methods based on feature subsequences [5, 28, 30], where some representative subsequences are selected from the complete sequence and are used as features for classification. Since anomaly prediction is regarded as a classification problem, early classification of time series has been the focus of research [5, 28–30, 42, 43].

Compared with classification methods based on global features, the classification method based on feature subsequences can effectively avoid irrelevant mode interference [19] and strong interpretability [20–22, 30]. However, the feature subsequence cannot be directly used for phase division, although it can characterize phases/stages. In the context of this paper, the feature-based subsequence method can be uniquely used as a sign of the evolution stage of anomalies.

In addition, recently developed deep learning (DL) for anomaly detection has shown strong learning ability with high classification accuracy [46], while the currently popular

hybrid deep learning-based anomaly detection techniques have proven to be effective in multiple tasks [47]. Since the hybrid models send extracted features to different anomaly detection methods, it is impossible to connect features directly to the representational learning in the hidden layers. An appropriate objective function and geometrical transformations have been proposed relatively by Ruff et al. [48] and Golan and El-Yaniv [49] to combine the encoding and detection steps for training a single neural model.

Moreover, existing anomaly detection techniques for COVID-19 data focus only on outbreak detection [50–53] in the COVID-19 tracking cases on the world wide. Many of these techniques use supervised machine learning, by assuming the existence of labeled training data. However, in real-world it is unavailable for new forms of outbreaks such as COVID-19.

To sum up, the existing anomaly prediction methods whether it is machine learning based or deep learning based, do not sufficiently emphasize the stages of abnormal prediction. Actually, taking stage as one of main considerations means that a new time series mining problem must be defined. Although the existing methods are not absolutely eligible for anomaly prediction, they are still of great reference value to the time series mining method based on feature subsequences and are more suitable for anomaly prediction.

3. Main Mathematical Background of Anomaly Prediction Using Time Series Mining

Definition 1 (univariate time series). **A** univariate time series indicates a set of real numbers that is recorded sequentially in ascending order of time stamp, such as $(t_1, x_1,..., t_n x_n)$, where $n \geq 1$, x_i is the observation value corresponding to the time stamp ti $(1 \leq i \leq n)$.

Definition 2 (multivariate time series). m $(m \geq 1)$ single-variable time series sharing a set of time stamps constitutes a multivariable time series, such as $(t_1, <x_{11},..., x_{m1}>)(t_n, <x_{1n},..., x_{mn}>)$, where $n \geq 1$, $(t_i, <x_{1i},..., x_{mi}>)$ are m observations corresponding to the timestamp t_i $(1 \leq i \leq n)$. m is called the dimension of the multivariate time series.

In the above definitions, for the sake of simplicity, the length of the time series is restricted to a finite value; actually, the length of the time series can be infinite, for instance, a sensor data stream in a long-term real-time update state can be considered as infinite time series. Moreover, a segment in a univariate time series is called a subsequence of the time series, which is precisely defined as follows:

Definition 3 (subsequence). Given a univariate time series $TS = (t_1, x_1), \cdots, (t_n, x_n), \forall (i, j), 1 \leq i \leq j \leq n$. $SS = (t_i, x_i), \cdots, (t_j, x_j)$ is a subsequence of TS. Mostly, time intervals of the time stamps in the time series are uniform. Under normal circumstances, it is to hide the time stamps and only preserve the sequence of observations.

4. Characteristics of Anomaly Prediction

Through the analysis of existing research results and the investigation of application fields, we hold the opinion that abnormal prediction should be characterized by:

4.1. Anomaly Detection/Prediction-Based Classification Problems. Roughly, existing anomaly detection algorithms are classified into two categories: "abnormal" refers to a pattern that deviates from "normal" [15, 18], and the so-called "normal" means that the data accords with the characteristics of most data [15] or accords with a certain hypothesis [18]; the other type regards anomaly detection as a classification problem of normal and abnormal samples [13, 44–49]. For the former, there are two problems: Firstly, it can be easily affected by interference waveforms. For example, "abnormal" changes will occur in certain monitoring indicators when touching patients' body in the intensive care unit, and medical staff does not need to intervene such changes. Secondly, it is impossible to specify the exact type of exception. As a result, users difficultly respond to anomalies properly. The above problems can be circumvented by the classification-based anomaly prediction method for the interference waveform; when just marking it as a normal mode, the classifier can identify the difference between it and the real abnormal waveform, thus avoiding mixing the two as much as possible. For anomaly types, when just assigning different types of anomalies to different class labels in the labeling process, the detection algorithm can be trained to automatically identify the type of anomalies.

4.2. Early Detection of Abnormal Signs with Continuous Monitoring. In the anomaly prediction problem, anomaly signs shall be captured. It is obvious that early detection of anomalies [29, 30] can provide users with more reaction time. However, only early detection of anomalies is not sufficient to complete the task of anomaly prediction. Firstly, for many anomalies, there is a phenomenon of "concept drift" [31] from the appearance of anomalies to the actual occurrence of anomalies, which is known as the evolution of anomalies. The evolution process can be divided into several stages, and there is different semantic information of abnormal signs in each stage. Lacking in-depth knowledge regarding the stage of the abnormal sign, users will difficultly take proper countermeasures. Secondly, even if users can be informed of the anomaly in a single warning, the domain knowledge and experience of users over the evolution process may not be sufficient to allow them to make effective response decisions at different stages. It is necessary for anomaly prediction algorithms to continually monitor the evolution process of anomaly signs, to inform users in time when the anomaly signs enter a new stage. Thirdly, if users can make targeted decisions at the current stage, they still need to evaluate whether their response measures are effective. This requires anomaly prediction algorithms to continue to monitor data. Based on the evolution process of the flow, we can judge whether the monitored object is recovered from the abnormal evolution state. Broadly, the anomaly prediction algorithm shall have the functions to

catch anomalies as early as possible, and stage and continuously monitor the evolution of anomaly signs.

4.3. Prioritization and Comprehensive Analysis of Multiple Data Streams for Anomaly Prediction. Most research on time series anomaly detection (such as [13, 18, 44]) lays emphasis on univariate time sequence. However, in a multifactors system, it is obvious that only a single variable cannot fully and accurately describe the behavior of the system. Now let us consider an air quality prediction problem. In the absence of consideration for other factors, the sudden increase in wind can make the air pollutants spread to the downwind area B from the heavily polluted area A at a higher speed, which may dramatically improve the air quality of the area B. Therefore, to accurately predict the air quality of place B, at least four indicators are needed as follows: including air quality of place A, (historical) air quality of place B, and wind direction and speed of place A. Obviously, this is a univariate time series mining problems that are unsolvable. We need to comprehensively analyze and predicate based on multiple data streams. Certainly, sometimes subject to conditions (for example, there is only one available monitoring indicator), we must make analysis based on a single data stream. Hence, methods should prioritize the comprehensive analysis of multiple data streams and support single data stream analysis. It is concluded that time series anomaly prediction shall be an early and staged classification process of multivariate time series.

5. Staged Abnormal Prediction of Time Series

On this issue, we shall understand the evolution process of anomalies based on the early capture of anomalies and continuous monitoring. To be more specific, we attempt to realize the automatic identification of each stage during the evolution process and predicate anomalies at each stage. In view of the novelty of this problem, a precise definition of this problem is mentioned herein. First of all, the concept of anormal antecedents of evolution shall be introduced.

Definition 4 (abnormal antecedents). Given a time series TS, its time span is $t_1,..., t_n$ ($n \geq 1$). Supposing that an abnormality occurs at time t_a, given a minimum efficient response time, Δt_{dcs} ($0 < \Delta t_{dcs} < t_a - t_1$) and the length of the predecessor Δt_{pfx} ($0 < \Delta t_{pfx} \leq t_a - \Delta t_{dcs} - t_1 + 1$), we call all the data in TS that fall within the time span $t_a - \Delta t_{dcs} - \Delta t_{pfx} + 1, \cdots, t_a - \Delta t_{dcs}$ as Δt_{pfx}-anomalous antecedents of TS.

The above definitions can be explained as follows. Firstly, they are to distinguish between the anomalous antecedents and the time series prefix. The latter is a sequence fragment consisting of the first several observations in a complete time series, while the former is a complete sequence that occurs before the anomaly. Secondly, the sequence TS in the above definitions can extend infinitely forward and backward along the time axis, respectively. As a matter of fact, in long-term online continuous monitoring, it is arduous for us to find the starting and ending points for the sequence; however, but considering the computational

and storage resources finiteness, the length of the antecedent Δt_{pfx} is used to "cut out" a finite section from the infinite data stream to carry out analysis. Thirdly, there must be a certain time interval between the last moment of abnormal antecedent and the moment of abnormal occurrence. On the one hand, the process of data sampling, transmission, and analysis of anomaly prediction algorithms will result in a certain delay; on the other hand, it is necessary to provide users with a certain response time to the warning, so as to make preparations for the exception. The sum of the above two delays is expressed as the minimum effective response time Δt_{pfx}. Fourthly, for the length of the antecedent Δt_{pfx}, it shall be sufficiently large. Specifically, it should cover the whole process of the evolution of abnormal signs in the data stream; meanwhile, we allow it to cover partial data of the evolution process of nonabnormal signs during the initial period. Furthermore, these parts will not cause significant interference. The reason is that the method based on feature subsequence can avoid interference from irrelevant patterns [19]. Now, we give the definition of time series phased anomaly prediction:

Definition 5 (time series phased abnormal prediction). The phased anomaly prediction of time series is divided into two parts, namely, offline training and online prediction. In the offline training module, the training set includes several m ($m \geq 1$) dimensional time sequence, and each sample has a class label c, $c \epsilon C = (c_0, c_1, \cdots, c_q)$. Among them, the sample with the class label c_0 is intercepted from the data stream that does not contain any abnormalities. Additionally, the samples with the class label c_i ($1 \leq i \leq q$) are, respectively, abnormal antecedents, and different class labels correspond to different types of anomalies. Given the minimum effective response time Δt_{dcs}, the training process aims to establish a rule set RS, where each rule R is a four-tuple ($\mathfrak{R}, \varepsilon_{mtc}, \varepsilon_{dis}, c_p$). Among them $\mathfrak{R}$, is called rule subject, ε_{mtc} is the matching condition for R, ε_{dis} is the matching release condition for R, and c_p is the class label corresponding to this rule. The body of the rule is as follows:

$$\left(\underset{1}{o}, \Delta t_1^{\min}, \Delta t_1^{\max}, \underset{1}{p} \right) \longrightarrow \cdots \longrightarrow \left(\underset{v}{o}, \Delta t_v^{\min}, \Delta t_v^{\max}, \underset{v}{p} \right). \quad (1)$$

Each four-tuple ($o, \Delta t^{\min}, \Delta t^{\max},$ and p)($1 \leq j \leq v$) in the body of the rule is called an evolution stage of rule R, where o is an abnormal sign, $\Delta t^{\min}$ and $\Delta t^{\max}$, respectively, indicate the shortest and longest time interval (estimated value) from the time when o is detected to the time of the anomaly occurrence, and p_j refers to the prediction confidence (estimated value). The main body of the rule is acquired by sorting out the evolution stages in chronological order.

In the online prediction module, given an m-dimensional time series TS to be classified, at any time t, for each rule in the rule set RS, $R = (\mathfrak{R}, \varepsilon_{mtc}, \varepsilon_{dis}, c_p)$, a binary indicator variable $z = f$ (TS, R, t) is used to indicate whether the

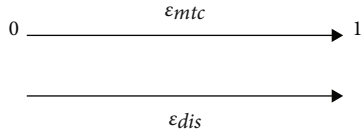FIGURE 1: An illustration of a finite state automata.

sequence TS matches the rule R at time t. The value of z can be represented by the finite state automata as Figure 1:

Among them, $z = 0$ means that the matching state is not established, and $z = 1$ means that the matching state is established. The above finite state automata are read as "When TS does not match R, if the matching condition ε_{mtc} is established, TS turns to match R; when TS matches R, if the matching cancellation condition ε_{dis} is established, TS turns to "Does not match R." If TS and R are matched, and an anomaly o_j is detected in it, according to the rule R, there is such prediction as "It is expected that an abnormality with class label c_p ($c_p = c_0$ means that there will be no abnormality in the short-term)" will occur within $[\Delta t^{min}, \Delta t^{max}]$ and the credibility of this prediction is p_j.

In the above definitions, the main body of the rule indicates the path of evolving anomalies, and each evolution stage corresponds to a certain anomaly sign. Several characteristic subsequences are used, which intensively appear in a shorter period. As a sign (Figure 2), each stage corresponds to a set of shortest and longest time intervals, which indicate the approximate time needed for evolving from the current stage to the occurrence of anomalies. For each stage, the confidence of its prediction shall be estimated. When dividing the rules in the training stage, the confidence threshold shall be set for the confidence of the stage; for each stage, the confidence threshold is unnecessarily same. Generally, it is expected that the first stage of the rule subject is highly credible. This is because the first observations of the abnormal antecedents do not correspond to the abnormal evolution process, and a higher confidence threshold can help ensure that the search of the first stage is really in the process of abnormal evolution; simultaneously, it is also expected that the various stages of the end of the evolution are highly credible. The reason is that at this stage, users shall formulate and take effective countermeasures within a short time, which relies on more reliable predictions; for other stages of the evolution process, the requirements for confidence can be less strict, but the lower bound of confidence shall also be set to prevent the introduction of irrelevant models.

Additionally, for each rule R, a matching condition and a matching cancellation condition are introduced. This is because the abnormal antecedents that have been matched may change the development path midway for some reasons and deviate from the original evolution process. For example, regarding critical patients, the condition that was gradually deteriorating may be relieved or even fully recovered due to active treatment; accordingly, it also allows the samples that originally deviated from the evolution process to return to the evolution process again, such as repeated illnesses. Certainly, this kind of regression to the evolution process may not continue to evolve from the next stage after the last match to the last stage but may "regress" to a certain stage before, "jump" after the certain stages, etc., as the case.

6. Challenges in Multiple Data Streams Analysis for Anomaly Prediction

6.1. Challenge 1: High Sampling Rate and Large Data Volume of Time Series Data Streams Severally Affect the Storage and Matching of Rules. In many cases, the sampling rate of time series data streams is very high, and massive data will be accumulated due to the long-term monitoring under high sampling rate conditions, which causes the following two problems. Firstly, a large amount of data may lead to an excessively large anomaly prediction rule set, which indirectly threatens the success of anomaly prediction. Although the length of a single feature subsequence may be limited, in long-term historical data, we may find many feature subsequences. Assuming these subsequences are directly stored, the size of the rule set used for abnormal prediction will be more than the size of the memory of storage capacity. With a high sampling rate, if data needs to be read from the disk during abnormal prediction, the rate of rule matching may not keep up with the speed of data update. As a result, abnormal prediction cannot be performed. Secondly, it takes too much time to carry out data processing using certain time-complex methods. For example, for two time series of length n, the time complexity of calculating the distance with the dynamic time warping measure is $O(n^2)$. Under circumstances that the data is not processed, this similarity measure with high time complexity may be hardly used in online real-time monitoring scenarios.

6.2. Challenge 2: The Time Bending Phenomenon in the Time Series Creates Obstacles to Similarity Matching. The value space of each observation in a time series is an infinite field of real number, which makes it almost impossible to have two time series (no matter how semantically similar they are) the same [23]. Unless the time series is discrete [18, 28], we can only carry out similarity matching (not exact match) for time series [24, 25, 32, 50]. However, there is a phenomenon called distortion in the time series, which poses a challenge to the similarity comparison between two time series [32]. Chen et al. [32] divided the bending into four situations: time shifting, time scaling, amplitude shifting, and amplitude scaling, among which the first two phenomena are collectively called time warp, and the latter two phenomena are collectively called amplitude warp. Table 1 shows the explanation of the above four bending phenomena, while Figure 3 shows the corresponding examples. Amplitude curvature can be handled by Z-normalization [25] (Figure 3), while time curvature often needs to be handled by a well-defined time series similarity measure [24, 25, 32, 50] to design such similarity. Measuring is not ordinary work.

6.3. Challenge 3: Dividing and Characterizing the Evolution Process of Anomalies Face Multiple Challenges

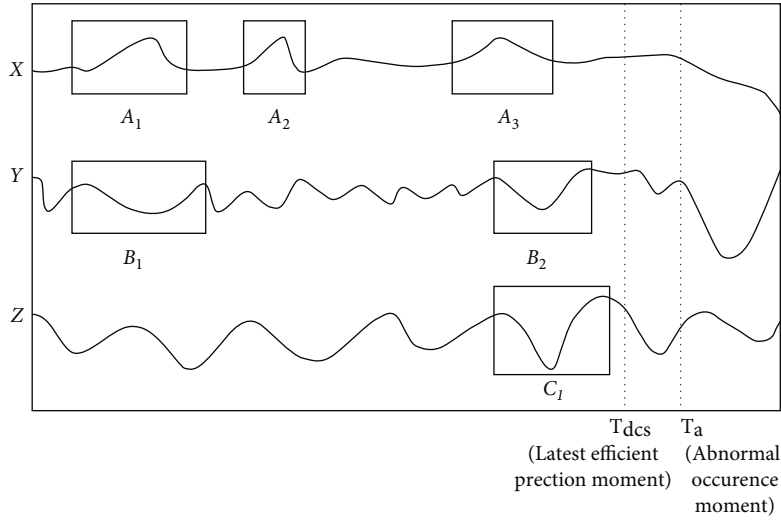*6.3.1. Challenge 3.1: The Evolution Process is Difficultly Segmented.* The time series has no natural segmentation

FIGURE 2: An example of phased anomaly prediction with feature subsequence set as a precursor. In this monitoring system based on a three-dimensional time series, anomalous antecedents are divided into three phases, which are characterized by feature subsequence sets {A1, B1},{A2}, and {A3, B2, C1}.

TABLE 1: Curving phenomena in time series.

| Types	Names	Meaning
Flection time space	Space-time dislocation	Similar waveforms in two sequences present different starting moments
	Different duration	The duration of similar waveforms in the two sequences is different
Amplitude bending	Amplitude misalignment	The balance positions of similar waveforms in the two sequences are different
Different amplitude	Fluctuations of similar waveforms in the two series	Different amplitude
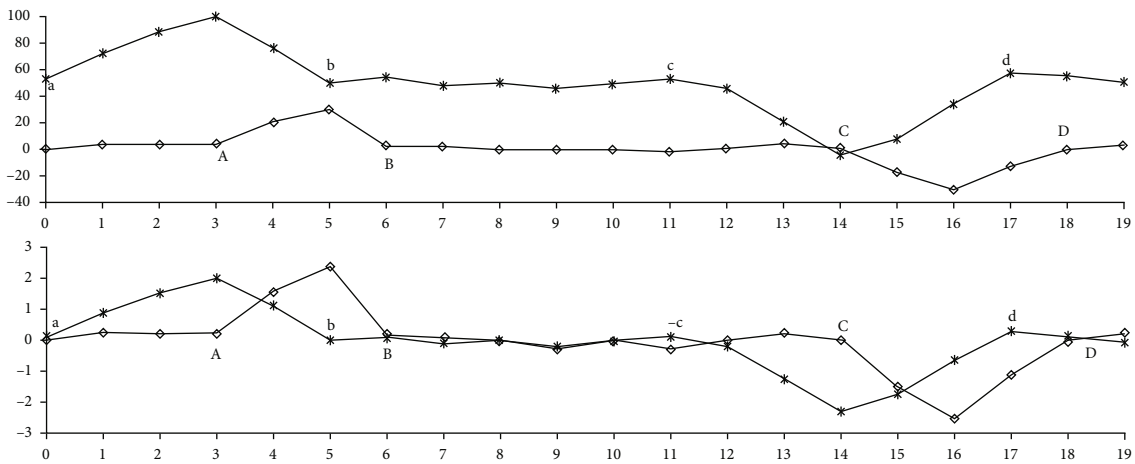
FIGURE 3: Curved phenomena in time series. Both the two sequences in the above figure contain 20 observations and have similar semantics: both of them consist of a crest (AB and ab) and a trough (AB and cd) that appear sometime after the crest appears. It should be noted that the two peaks and two troughs appear at different moments, which is called a time dislocation; the time spans of the two peaks and two troughs are different, which is a phenomenon of different duration; the balance positions of the two sequences are different, which is a phenomenon of amplitude dislocation. The fluctuation amplitudes of the two peaks and two troughs are different, which is a phenomenon of different amplitudes. The two sequences in Figure 1 are resulted from Z-standardization of the sequence in this figure, in which amplitude curvature is effectively alleviated, while time curvature is on the contrary.

[23], so it is possible to accurately match the changes in the potential semantic information of time series (i.e., concept drift [31]). The segmentation is difficult. In the context of anomaly prediction, we attempt to segment the evolution process of anomaly signs, and the latent semantic information of this process is more profound. Limited by domain knowledge, the evolution of anomalies may even be difficult to describe in natural language at this stage, which probably pose more challenges to segmentation.

6.3.2. Challenge 3.2: The Feature Subsequences for Each Segment are Difficultly Found. We make efforts to characterize each stage with feature subsequences. However, it is not easy to train the feature subsequences separately for the data of each class label on a given segment. On the one hand, as mentioned in Challenge 1, we can hardly match the time series accurately. Therefore, for the (candidate) feature subsequence, a threshold distance must be found for it [20–22, 32]. When the distance between a certain subsequence and it is less than this threshold, we deem that the feature subsequence is matched, and setting this A threshold is not a trivial task; on the other hand, an evaluation index shall be designed to evaluate the feature significance of candidate subsequences, and the many factors shall be taken into account.

7. Outlook and Future Work

To solve the brand-new problem of time series anomaly prediction, the proposed method is to achieve offline rule discovery and online anomaly prediction. In the offline rule discovery stage, the rules described in Definition 5 are discovered from historical data, and the generated rules are stored in the rule database; in the online anomaly prediction stage, the monitored data stream is matched with the rules in the rule database to predict anomalies; for anomalies that are not successfully predicted, the corresponding data is entered into the offline rule discovery process to update the rule database for better prediction of abnormalities in the future. This practice aims to roughly divide data stream of the original time series using preprocessing module and design the phased abnormal prediction model. The research objectives of preprocessing module are to design an effective time series compression method and to design an effective time series similarity measure based on the compression method.

7.1. An Effective Time Series Compression Model. The anomaly prediction process shall be expanded on the compressed data stream, which is largely resulted from the efficiency considerations of storage and data processing. In our previous research work [12, 15, 16], we continued to focus on two types of time series compression algorithms, DP [51] and PLR [26, 52]. DP is a compression method based on key points, following the principle of divide and conquer: Given a time series, the process of DP compression is as follows:

(1) The first dimension and the last dimension of the sequence is taken as anchor (positioning) and floating points, respectively

(2) The maximum vertical distance between the anchor point and floating point is found, and the point where the distance is greater than a given threshold is the cut point

(3) The tangent point is taken as the new floating point of the front section and the positioning point of the back section

(4) If the cut point is not found, stop the algorithm

(5) The algorithm runs iteratively on two cut-out segments

PLR is a common piecewise linear compression method. Given a segment, PLR uses the least square method to linearly fit the observations of each segment. Practically, the segmentation and fitting process of PLR is usually carried out iteratively to find the best segmentation.

The above two methods have their respective advantages and disadvantages: specifically speaking, DP can effectively identify the local key points of time series, while the effective information contained in the points outside the key points is not utilized; the effect of PLR generally relies on the segmentation process, and finding effective segmentation is not trivial work.

By combining DP's ability to identify key points and excellent fitting ability of PLR, this paper is to design a new compression method. These key points identified by DP method are regarded as the endpoints of each segment, and then linear fitting on each segment is performed. This method is called DP-PLR. To preliminarily verify the feasibility of this idea, a segment of electrocardiogram data (a segment of ECG pictures) that belongs to the first patient in the MGH/MF waveform database [53–55] was intercepted. After *Z*-normalization was used, it was compressed with DP and DP-PLR, respectively. The results are shown in Figure 4.

In order to quantitatively evaluate the degree of preservation of the original semantic information by compression, the reconstruction error of the sum of squares from the compressed sequence to the original sequence is calculated: given an original sequence $\bar{Y} = (y_1, \cdots, y_n)$ and compressed, $\bar{Y} = (\bar{y}_1, \cdots, \bar{y}_n)$, the reconstruction error of the two is defined as $Err(Y, \bar{Y}) = (y_1 - \bar{y}_1)^2 + \cdots + (y_n - \bar{y}_n)^2$.

Note that DP is a sampling method, which means that the above reconstruction error formula cannot be directly used to evaluate DP. To this end, we connect each pair of adjacent key points in DP with a straight line to complement the missing values. After calculation, the reconstruction error of DP is 28.4819, and the reconstruction error of DP-PLR is 12.9157, which indicates that linear fitting can maintain the original semantic information.

Furthermore, the termination condition of DP can also be adjusted. It should be noted that a vertical distance threshold is used as the termination condition in the previous DP, and it is difficult to set this threshold. A greedy algorithm is proposed to determine whether to continue segmentation. Each segmentation is expected to reduce
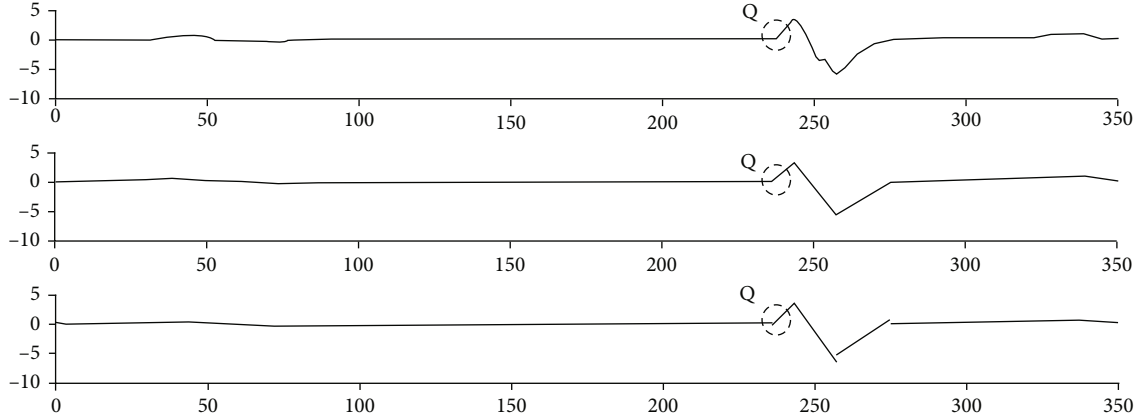
FIGURE 4: Example of time series compression. From top to bottom are the Z-standardized ECG sequence, the DP compressed sequence, and the DP-PLR compressed sequence. The reconstruction errors of the latter two are 28.4819 and 12.9157, respectively.

the reconstruction error. Thus, if the reconstruction error is reduced after a certain segmentation, then the segmentation is continued; otherwise, the segmentation stops.

7.2. How to Design Similarity Measures for Compressed Time Series.

One of the basic techniques of time series mining is the similarity matching of time series [24, 25, 32, 37, 50, 56–72], while the advantages and disadvantages of similarity measures directly affect the performance of similarity matching. How to design similarity measures for compressed time series is the second important problem to be solved in the processing stage. Mostly, the existing similarity measures [24, 25, 32, 50] focus on comparing the similarity of time series in the original space, instead of the compressed series. However, this does not affect our design ideas for drawing on existing methods.

In Section 6, we suggest that time warping is one of the important factors affecting the design of similarity measures. As the best distance measure in time series mining problems, the dynamic time warping (DTW) [25, 50] measure specifically addresses this problem [25].

DTW is based on such an inherent logic: time warping means the correspondence between the timestamps of the time series and the semantic correspondence are inconsistent. DTW uses the idea of dynamic programming, aiming to adjust the correspondence of the timestamps at the time of ensuring the timing relationship. The relationship makes it move along the direction of semantic correspondence, in order that the global semantic gap between the two sequences is narrowed sufficiently. This process is called the "alignment" of the two sequences.

Specifically, given two time series $X = x_1, \cdots, x_m$ and $Y = y_1, \cdots, y_n$, DTW establishes a distance matrix of $m \times n$. The value of the position (i, j) in the matrix is $dist(i, j) = (x_i - y_j)^2$. According to this distance matrix, the distance of the subsequence composed of the first i values of X and the first j values of Y can be calculated by the following recursive formula: $DTW(i, j) = dist(i, j) + \min(DTW(i, j - 1), DTW(i - 1, j), DTW(i - 1, j - 1))$,

Figure 5 shows the correspondence between the data points in the two sequences (after Z-standardization) in Figure 2 under DTW, and the data points in the peaks and troughs of the two basic correspondences correctly. It should be noted that, in order to prevent overfitting, a "curved window"[67] (shown as in Figure 6) shall be generally added to the DTW to limit the regular path.

Under DP-PLR compression, we intend to carry out similarity comparison using logic similar to DTW. However, in this case, it is not the correspondence between data points that needs to be regularized, but the correspondence between segments. Specifically, given two compressed sequences, we establish the distance matrix between the segments, and then find the mapping relationship between the segments by a method similar to DTW.

7.3. An Efficient Anomaly Prediction Algorithm with Phase Division.

An efficient anomaly prediction algorithm mainly includes the following two modules.

7.3.1. Design of Phase Division Model for Efficient Anomaly Prediction.

In online prediction, we use feature subsequences to characterize each stage, and different stages are likely to have different feature subsequences. We must first perform a preliminary segmentation of the training data, look for feature subsequences on each segment, further adjust the segmentation according to the characterization ability of the feature subsequences, and then iteratively carry out the process of "segmentation-feature subsequence recognition" until enough good stage identification and characterization performance to be achieved. It is proposed to learn from a dynamic segmentation method [63] in the time series index [24, 63].

For simplicity, it is required that all sequences in the training set have the same length. Specifically, we divide the stages according to the following method:

(1) Firstly, each sample in the training set is evenly divided into several segments, and the number of segments of all samples is the same
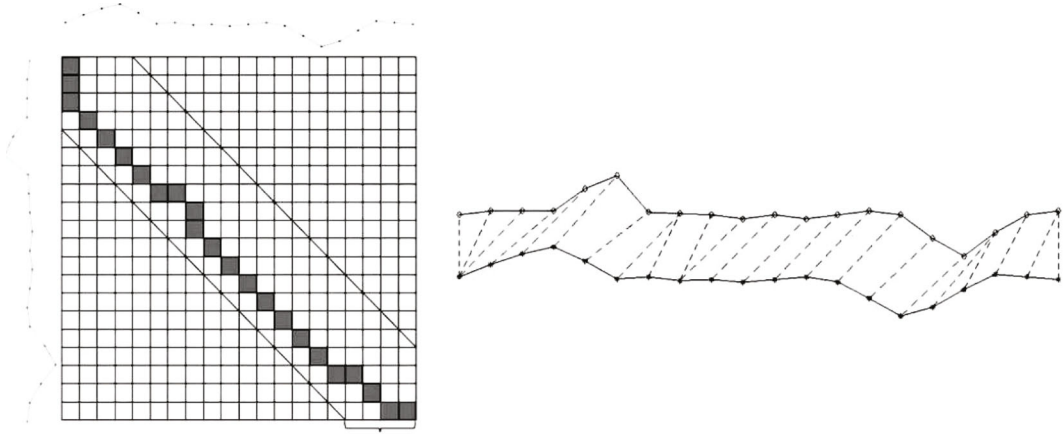
FIGURE 5: DTW diagram. The left picture shows the distance matrix and the regularization path, and the right picture shows the correspondence between the data points in the two sequences after regularization.
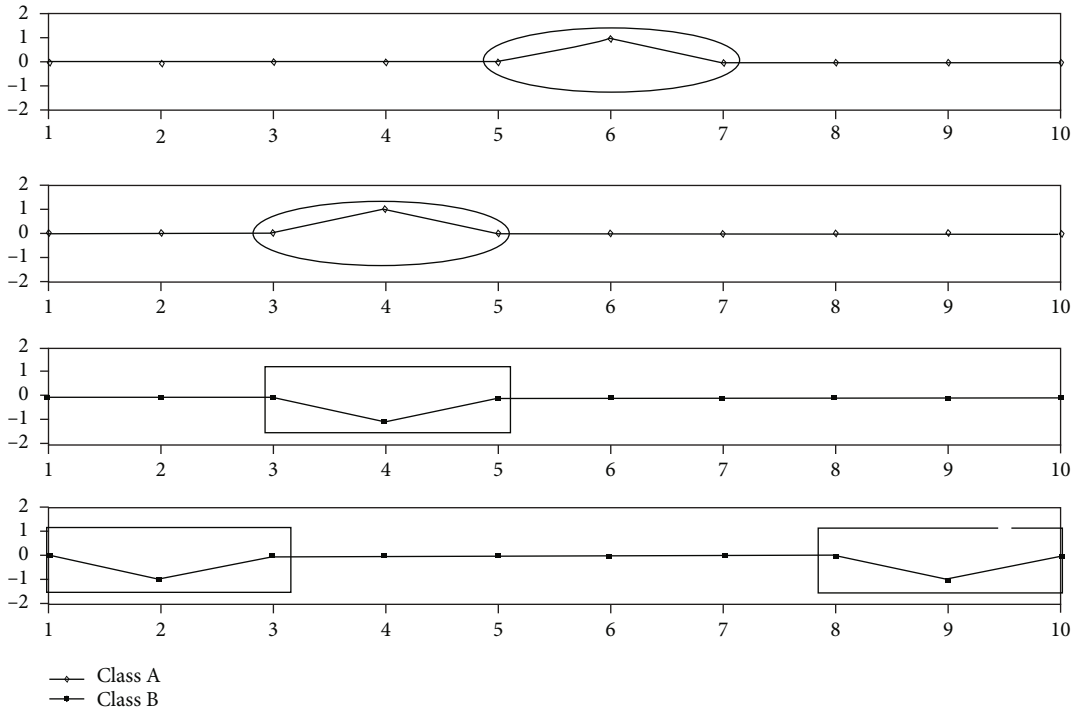

FIGURE 6: Illustration of shapelets. The sequence in the figure is divided into A and B. The subsequence circled by the ellipse is the shapelets of feature A, while the rectangles are the shapelets of feature B. This figure is adapted from Figure 2 of [30].

(2) Find a set of characteristic subsequences on each segment (As detailed in Section 7.2) as the characterization of the current segment

(3) For each segment, score the characterization ability of the segment in its feature subsequence set, assuming that the score of a certain segment seg is Q_{seg} (As detailed in Section 7.2)

(4) For each segment, using its midpoint as the boundary, perform feature subsequence identification and scoring on the left and right segments, respectively, assuming that the scores of the left and right segments of seg are Q_{seg}^L and Q_{seg}^R, respectively, when

$Q_{seg}^L + Q_{seg}^R \geq 2Q_{seg}$, replace seg with left and right ends

(5) Repeat the three processes of (2), (3), and (4) and stop until a certain termination condition (such as the minimum distance between the feature subsequences of two adjacent segments is less than a threshold) is satisfied

7.3.2. Efficient Construction of Feature Subsequence Set. Each stage of the evolution process of abnormal signs includes a feature subsequence set. Specifically, for the training set with the number of class labels n_c and the sequence dimension m,
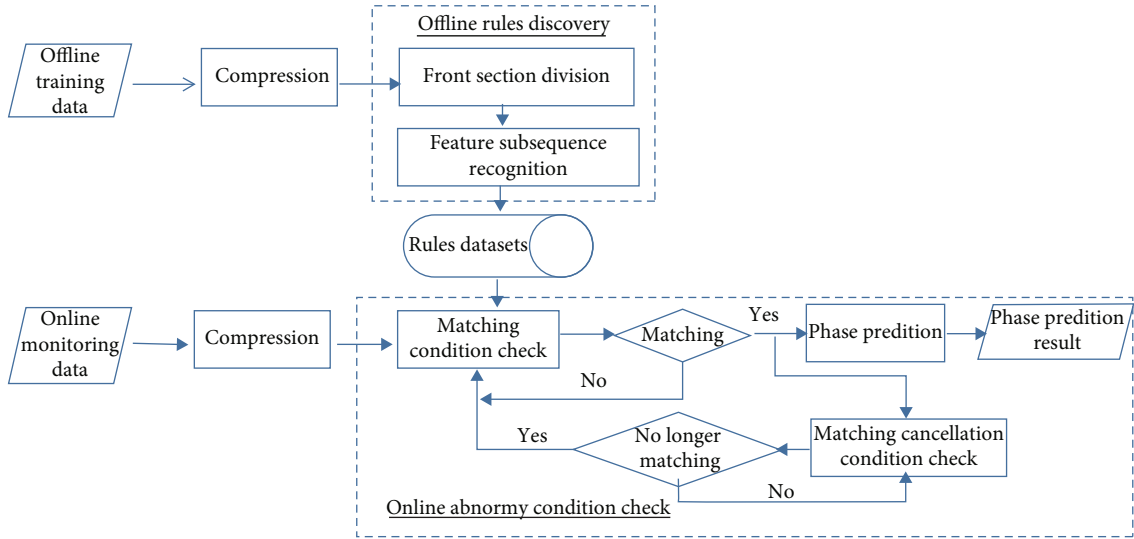
FIGURE 7: Technical framework.

in the iterative segmentation process described above, the feature subsequence set *FS* satisfies the following conditions: each class label is characterized by the ms $(1 \leq ms \leq m)$ subsequence in *FS*, which is indispensable, and the above subsequences must come from different dimensions. In other words, for each class label, we use subsequences from one or more dimensions to characterize, and at most one subsequence is extracted from each dimension. For this reason, the selection of feature subsequences needs to go through two subprocesses: firstly, extract a few feature subsequences that can characterize this dimension on each dimension to construct a "subsequence pool"; and secondly, select several subsequences from different dimensions in the pool, so as to construct a set of characteristic subsequences.

The construction of the subsequence pool is considered firstly. We regard abnormal prediction as a classification problem, and a common feature subsequence in time series classification is time series shapelet(s) [20–22, 30] (hereinafter referred to as shapelet(s)), which can effectively represent and distinguish the sequence fragments of various characteristics (Figure 6). Specifically, given a shapelet and a corresponding distance threshold δ, we call a certain sequence to include the shapelet, if and only if the distance between at least one subsequence in the sequence and the shapelet shall be less than δ [20–22].

Subsequently, it is necessary to screen the combination of subsequences from the subsequence pool to obtain the final set of characteristic subsequences. We use a scoring function to score each candidate as follows:

(1) The confidence of the candidate. Namely, how can the candidate represent its class label. The classic definition of confidence [28] is adopted. For the number of samples containing the candidate and the class label corresponding to the candidate's class label, and the ratio of the number of samples containing the candidate as the confidence score of the candidate item, this score is set to Q_{conf}

(2) The synchronization of each sequence in the candidates. In a candidate item, the time may not be entirely consistent, when the subsequences from different dimensions appear. It is expected that the time interval between their appearances is short enough. For this reason, we can use the first and last occurrences of the two subsequences in the candidate item. The reciprocal of the time interval is used as the synchronicity score, and this score is set to Q_{diff}. The final score can be calculated using the equation: $Q = Q_{conf} \times Q_{diff}{}^{\alpha}$. Among them, α is used to balance the weight between the above two. We choose the candidate with the highest final score as the feature subsequence set

(3) In summary, the technical framework of the proposed anomaly prediction algorithm has been illustrated in Figure 7, which can be roughly divided into two stages, namely, discrete rule discovery and online anomaly prediction. Both training and monitored data are first compressed with the methods based on DP [51] and PLR [26, 52]. In the discrete rule discovery stage, the dynamic segmentation [63] method is applied to divide the training data in stages, and find the feature subsequence set on each segment by the method described in Section 5. The process of feature subsequence set construction is organically integrated into the process of stage division (Section 5), and iterative segmentation is achieved until the desired segmentation result is achieved. According to the final segmentation results, the shortest and longest time for the occurrence of anomalies at each stage (characterized by several feature subsequences) and the prediction credibility are estimated. In addition, rules matching and matching contact conditions need to be specified, thereby generating rules. The generated rules are entered into the rules database. At the online

anomaly monitoring stage, the monitored data is matched with the rules in the rule database in real time. Once it meets the matching conditions of a certain rule, the evolution process of anomaly warnings will be monitored: each monitoring stage (namely, the feature subsequence set corresponding to this stage is matched), and the prediction result corresponding to this stage are made pursuant to the rules. While making phased predictions, whether the data stream deviates from the matched rules shall also be continuously monitored. Once the matching cancellation condition is established, the data stream will not be considered to be consistent with the previously matched rules anymore.

8. Conclusion

Data stream mining strategies has emerged as mass data can hardly be produced and sorted. Even though there are many techniques, this research area still lacks in approaches to mine data streams composed of multiple time series, which has applications in financial, medical, and environmental monitoring.

Time-series comprise a large subset of most streaming data, which pose a serious challenge to machine learning applications. This is because there is generally a high order of cardinality. Taking the industrial sensor as an example if you want to predict machine failure or maintenance requirements, you might have to deal with thousands of sensors, each of which (at each industrial facility) may have its own set of time stamps. This means that many thousands of individual time series require a machine learning model to be trained and launched in production under normal circumstance. In terms of data streams mining, it is most difficult to deal with concept drifts that toughen similarity matching and measurement.

Time series anomaly prediction is schema adaptive, which means that you can derive insights without any data preparation. In this way, a variety of data sources can be explored, compared, and correlated easily. Additionally, such prediction provides SQL-like filters and aggregates, which are used to construct, visualize, compare, and overlay various time series patterns, and save and share queries.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] G. Huang, J. He, J. Cao, Z. Qiao, M. Steyn, and K. Taraporewalla, *A Real-Time Abnormality Detection System for Intensive Care Management*, ICDE, 2013.

[2] S. Yousefzadeh-Chabok, F. Ranjbar-Taklimie, R. Malekpouri, and A. Razzaghi, "A time series model for assessing the trend and forecasting the road traffic accident mortality," *Archives of Trauma Research*, vol. 5, no. 3, article e36570, 2016.

[3] C.-C. Chia and Z. Syed, *Scalable noise mining in long-term electrocardiographic time-series to predict death following heart attacks*, KDD, 2014.

[4] I. T. Telci, K. Nam, J. Guan, and M. M. Aral, "Optimal water quality monitoring network design for river systems," *The Journal of Environmental Management*, vol. 90, no. 10, pp. 2987–2998, 2009.

[5] A. McGovern, D. H. Rosendahl, R. A. Brown, and K. K. Droegemeier, "Identifying predictive multi-dimensional time series motifs: an application to severe weather prediction," *Data Mining and Knowledge Discovery*, vol. 22, no. 1-2, pp. 232–258, 2011.

[6] Y. Zhu and D. Shasha, "StatStream: statistical monitoring of thousands of data streams in real time," in *Proceedings of the 28th International Conference on Very Large Data Bases (VLDB'02). VLDB Endowment*, pp. 358–369, Hongkong, China, 2002.

[7] J. Chen, D. J. DeWitt, F. Tian, and Y. Wang, "NiagaraCQ: a scalable continuous query system for internet databases," in *Proceedings of ACM SIGMOD International Conference on Management of Data (SIGMOD'00)*, pp. 379–390, Dallas Texas USA, 2000.

[8] D. A. Davis, N. V. Chawla, N. Blumm, N. A. Christakis, and A.-L. Barabási, *Predicting individual disease risk based on medical history*, CIKM, 2008.

[9] E. Horvitz, "From data to predictions and decisions: enabling evidence-based healthcare," *Computing Community Consortium*, vol. 6, 2010.

[10] J. Brotzge and S. Erickson, "NWS tornado warnings with zero or negative lead times," *Weather and Forecasting*, vol. 24, no. 1, pp. 140–154, 2009.

[11] K. M. Simmons and D. Sutter, "Tornado warnings, lead times, and tornado casualties: an empirical investigation," *Weather and Forecasting*, vol. 23, no. 2, pp. 246–258, 2008.

[12] J. He, Y. Zhang, G. Huang, and P. de Souza, "CIRCE: correcting imprecise readings and compressing excrescent points for querying common patterns in uncertain sensor streams," *Information Systems*, vol. 38, no. 8, pp. 1234–1251, 2013.

[13] J. Gudmundsson, M. van Kreveld, and B. Speckmann, "Efficient detection of patterns in 2D trajectories of moving points," *GeoInformatica*, vol. 11, no. 2, pp. 195–215, 2007.

[14] J. Zhou, S. Zhu, X. Huang, and Y. Zhang, "Enhancing time series clustering by incorporating multiple distance measures with semi-supervised learning," *Journal of Computer Science and Technology*, vol. 30, no. 4, pp. 859–873, 2015.

[15] A. Al-Shammari, R. Zhou, M. Naseriparsaa, and C. Liu, "An effective density-based clustering and dynamic maintenance framework for evolving medical data streams," *International Journal of Medical Informatics*, vol. 126, pp. 176–186, 2019.

[16] G. Huang, Y. Zhang, J. He, and Z. Ding, "Efficiently retrieving longest common route patterns of moving objects by summarizing turning regions," Tech. Rep. 1, PAKDD, 2011.

[17] Z. Qiao, G. Huang, J. He et al., "Discovering semantics from multiple correlated time series stream," Tech. Rep. 2, PAKDD, 2013.

[18] E. J. Keogh, J. Lin, and A. W.-C. Fu, *HOT SAX: Efficiently Finding the Most Unusual Time Series Subsequence*, ICDM, 2005.

[19] B. Hu, Y. Chen, and E. J. Keogh, *Time Series Classification under More Realistic Assumptions*, SDM, 2013.

[20] L. Ye and E. J. Keogh, *Time Series Shapelets: A New Primitive for Data Mining*, KDD, 2009.

[21] A. Mueen, E. J. Keogh, and N. E. Young, *Logical-Shapelets: An Expressive Primitive for Time Series Classification*, KDD, 2011.

[22] T. Rakthanmanon and E. J. Keogh, "Fast shapelets: a scalable algorithm for discovering time series shapelets," SDM, 2013.

[23] M. Shokoohi-Yekta, Y. Chen, B. J. L. Campana, B. Hu, J. Zakaria, and E. J. Keogh, *Discovery of Meaningful Rules in Time Series*, KDD, 2015.

[24] H. Ding, G. Trajcevski, P. Scheuermann, X. Wang, and E. J. Keogh, "Querying and mining of time series data: experimental comparison of representations and distance measures," *PVLDB*, vol. 1, no. 2, pp. 1542–1552, 2008.

[25] T. Rakthanmanon, B. J. L. Campana, A. Mueen et al., *Searching and Mining Trillions of Time Series Subsequences under Dynamic Time Warping*, KDD, 2012.

[26] E. J. Keogh, *Fast Similarity Search in the Presence of Longitudinal Scaling in Time Series Databases*, ICTAI, 1997.

[27] L. Ulanova, T. Yan, H. Chen, G. Jiang, E. J. Keogh, and K. Zhang, *Efficient Long-Term Degradation Profiling in Time Series for Complex Physical Systems*, KDD, 2015.

[28] Z. Xing, J. Pei, G. Dong, and P. S. Yu, *Mining Sequence Classifiers for Early Prediction*, SDM, 2008.

[29] Z. Xing, J. Pei, and P. S. Yu, *Early Prediction on Time Series: A Nearest Neighbor Approach*, IJCAI, 2009.

[30] Z. Xing, J. Pei, P. S. Yu, and K. Wang, "Extracting interpretable features for early classification on time Series," SDM, 2011.

[31] J. Zhou and A. K. H. Tung, *SMiLer: A Semi-Lazy Time Series Prediction System for Sensors*, SIGMOD conference, 2015.

[32] Y. Chen, M. A. Nascimento, B. Chin Ooi, and A. K. H. Tung, *SpADe: On Shape-based Pattern Detection in Streaming Time Series*, ICDE, 2007.

[33] X. X. Yin, Y. Zhang, J. Cao, J. L. Wu, and S. Hadjiloucas, "Exploring the complementarity of THz pulse imaging and DCE-MRIs: toward a unified multi-channel classification and a deep learning framework," *Computer Methods and Programs in Biomedicine*, vol. 137, pp. 87–114, 2016.

[34] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," in *Proceedings of the 8th ACM SIGMOD workshop on research issues in data mining and knowledge discovery. ACM*, pp. 2–11, San Diego California, 2003.

[35] A. Bagnall, J. Lines, J. Hills, and A. Bostrom, "Time-series classification with cote: the collective of transformation-based ensembles," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 9, pp. 2522–2535, 2015.

[36] X. X. Yin, S. Hadjiloucas, J. He, Y. Zhang, Y. Wang, and D. Zhang, "Application of complex extreme learning machine to multiclass classification problems with high dimensionality: A THz spectra classification problem," *Digital Signal Processing*, vol. 40, pp. 40–52, 2015.

[37] Y. Wang, L. Sun, and S. Subramani, "CAB: classifying arrhythmias based on imbalanced sensor data," *KSII Transactions on Internet & Information Systems.*, vol. 15, no. 7, pp. 2304–2320, 2021.

[38] L. Sun, Q. Yu, D. Peng, S. Subramani, and X. Wang, "Fogmed: a fog-based framework for disease prognosis based medical sensor data streams," *Computers, Materials & Continua*, vol. 66, no. 1, pp. 603–619, 2021.

[39] X. X. Yin, B. W. H. Ng, B. Ferguson, and D. Abbott, "Wavelet based local tomographic image using terahertz techniques," *Digital Signal Processing*, vol. 19, no. 4, pp. 750–763, 2009.

[40] A. Pandya, O. Odunsi, C. Liu, A. Cuzzocrea, and J. Wang, "Adaptive and efficient streaming time series forecasting with Lambda architecture and spark," in *2020 IEEE International Conference on Big Data (Big Data)*, pp. 5182–5190, Atlanta, GA, USA, 2020.

[41] X. X. Yin, B. W. H. Ng, B. Ferguson, S. P. Mickan, and D. Abbott, "2-D wavelet segmentation in 3-D T-ray tomography," *IEEE Sensors Journal*, vol. 7, no. 3, pp. 342-343, 2007.

[42] J. J. Rodríguez, C. J. Alonso, and H. Boström, "Boosting interval based literals," *Intelligent Data Analysis*, vol. 5, no. 3, pp. 245–262, 2001.

[43] A. Bregón, M. A. S. Hurtado, J. J. Rodríguez, C. J. Alonso, B. P. Junquera, and Q. I. Moro, *Early Fault Classification in Dynamic Systems Using Case-Based Reasoning*, CAEPIA, 2005.

[44] Y. Chen, E. Keogh, B. Hu et al., "The UCR time series classification archive," 2015, http://www.cs.ucr.edu/~eamonn/time_series_data/.

[45] T. Reutterer, A. Mild, M. Natter, and A. Taudes, "A dynamic segmentation approach for targeting and customizing direct marketing campaigns," *Journal of Interactive Marketing*, vol. 20, no. 3-4, pp. 43–57, 2006.

[46] Z. Ji, J. Gong, and J. Feng, "A novel deep learning approach for anomaly detection of time series data," *Scientific Programming*, vol. 2021, Article ID 6636270, 11 pages, 2021.

[47] N. Davis, G. Raina, and K. Jagannathan, "A framework for end-to-end deep learning-based anomaly detection in transportation networks," *Transportation Research Interdisciplinary Perspectives*, vol. 5, p. 100112, 2020.

[48] L. Ruff, N. Görnitz, L. Deecke et al., "Deep one-class classification," in *Proceedings of the International Conference on Machine Learning*, pp. 4390–4399, Stockholm Sweden, 2018.

[49] I. Golan and R. El-Yaniv, "Deep anomaly detection using geometric transformations," in *32nd Conference on Neural Information Processing Systems (NeurIPS 2018)*, pp. 9758–9769, Montréal, Canada, 2018.

[50] Y. Karadayi, M. N. Aydin, and A. S. Ogrenci, "Unsupervised anomaly detection in multivariate spatio-temporal data using deep learning: early detection of COVID-19 outbreak in Italy," *IEEE Access*, vol. 8, pp. 164155–164177, 2020.

[51] T. Jombart, S. Ghozzi, D. Schumacher et al., "Real-time monitoring of COVID-19 dynamics using automated trend fitting and anomaly detection," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 376, no. 1829, article 20200266, 2021.

[52] G. Zhu, J. Li, Z. Meng et al., "Learning from large-scale wearable device data for predicting the epidemic trend of COVID-19," *Discrete Dynamics in Nature and Society*, vol. 2020, Article ID 6152041, 8 pages, 2020.

[53] I. E. Agbehadji, B. O. Awuzie, A. B. Ngowi, and R. C. Millham, "Review of big data analytics, artificial intelligence and nature-inspired computing models towards accurate detection of COVID-19 pandemic cases and contact tracing," *International Journal of Environmental Research and Public Health*, vol. 17, no. 15, p. 5330, 2020.

[54] Y. Alexander and R. P. W. Duin, "Novelty detection using self-organizing maps," *Progress Connect.-Based Info. Syst*, vol. 2, pp. 1322–1325, 1997.

[55] V. Vercruyssen, W. Meert, G. Verbruggen, K. Maes, R. Bäumer, and J. Davis, "Semi-supervised anomaly detection with an application to water analytics," in *2018 IEEE International Conference on Data Mining (ICDM)*, pp. 527–536, Singapore, 2018.

[56] G. Huang, Y. Zhang, J. Cao, M. Steyn, and K. Taraporewalla, "Online mining abnormal period patterns from multiple medical sensor data streams," *World Wide Web*, vol. 17, no. 4, pp. 569–587, 2014.

[57] F. Liu, X. Zhou, J. Cao et al., "Anomaly detection in quasi-periodic time series based on automatic data segmentation and attentional LSTM-CNN," *IEEE Transactions on Knowledge and Data Engineering*, p. 1, 2020.

[58] J. Rogala, E. Kublik, R. Krauz, and A. Wróbel, "Resting-state EEG activity predicts frontoparietal network reconfiguration and improved attentional performance," *Scientific Reports*, vol. 10, no. 1, p. 5064, 2020.

[59] J. Donald and J. Berndt, *Clifford: Using Dynamic Time Warping to Find Patterns in Time Series*, KDD Workshop, 1994.

[60] H. David and T. K. Peucker, "Algorithms for the reduction of the number of points required to represent a digitized line or its caricature," *Cartographica*, vol. 10, no. 2, pp. 112–122, 1973.

[61] A. L. Goldberger, L. Amaral, L. Glass et al., "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[62] S. K. Khare, V. Bajaj, and U. R. Acharya, "SPWVD-CNN for automated detection of schizophrenia patients using EEG signals,," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, 2021.

[63] J. P. Welch, P. J. Ford, R. S. Teplick, and R. M. Rubsamen, "The Massachusetts General Hospital-Marquette Foundation hemodynamic and electrocardiographic database – comprehensive collection of critical care waveforms," *Journal of Clinical Monitoring*, vol. 7, no. 1, pp. 96-97, 1991.

[64] "The MGH/MF Waveform Database," October 2021, https://physionet.org/content/mghdb/1.0.0/.

[65] M. N. Nguyen, X. L. Li, and S. K. Ng, "Ensemble based positive unlabeled learning for time series classification," in *Database Systems for Advanced Applications. DASFAA 2012*, S. Lee, Z. Peng, X. Zhou, Y. S. Moon, R. Unland, and J. Yoo, Eds., vol. 7238 of Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2012.

[66] G. He, Y. Duan, Y. Li, T. Qian, J. He, and X. Jia, "Active learning for multivariate time series classification with positive unlabeled data," in *2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 178–185, Vietri sul Mare, Italy, November 2015.

[67] X. Qiu, S. Liang, and Y. Zhang, "Simultaneous ECG heartbeat segmentation and classification with feature fusion and long term context dependencies," in *Advances in Knowledge Discovery and Data Mining*, vol. 12085, p. 371, Springer, 2020.

[68] X. Qiu, S. Liang, L. Meng, Y. Zhang, and F. Liu, "Exploiting feature fusion and long-term context dependencies for simultaneous ECG heartbeat segmentation and classification," *International Journal of Data Science and Analytics*, vol. 11, no. 3, pp. 181–193, 2021.

[69] C. Gómez, P. Arbeláez, M. Navarrete, C. Alvarado-Rojas, M. le van Quyen, and M. Valderrama, "Automatic seizure detection based on imaged-EEG signals through fully convolutional networks," *Scientific Reports*, vol. 10, no. 1, article 21833, 2020.

[70] S. L. Oh, J. Vicnesh, E. J. Ciaccio, R. Yuvaraj, and U. R. Acharya, "Deep convolutional neural network model for automated diagnosis of schizophrenia using EEG signals," *Applied Sciences*, vol. 9, no. 14, p. 2870, 2019.

[71] J. He, J. Rong, L. Sun, H. Wang, and Y. Zhang, "An advanced two-step DNN-based framework for arrhythmia detection," in *Advances in Knowledge Discovery and Data Mining*, vol. 12085, pp. 422–434, Springer, 2020.

[72] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *Signal Processing*, vol. 26, no. 1, pp. 43–49, 1978.