

Abnormal pattern recognition for online inspection in manufacturing process based on multi-scale time series classification

Xiangyu Bao¹, Yu Zheng^{1*}, Liang Chen¹, Dianliang Wu¹, Xiaobo Chen¹, Ying Liu²

¹*School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China.

² School of Engineering, Cardiff University, Cardiff, CF10 3AT, UK.

*Corresponding author(s). E-mail(s): yuzheng@sjtu.edu.cn;

Abstract

The collection of large volumes of temporal data during the production process is streamlined in a cyber manufacturing environment. The ineluctable abnormal patterns in these time series often serve as indicators of potential manufacturing faults. Consequently, the presence of effective analytical methods becomes essential for monitoring and recognizing these abnormal manufacturing patterns. However, the extensive process data may contain various minor abnormal patterns, typically reflecting changes in production status influenced by multiple anomalous causes. This study introduces an approach for recognizing abnormal manufacturing patterns through multi-scale time series classification (TSC). Long-term process signals undergo slicing using dynamically sized observation windows and subsequent classification at multiple scales employing our proposed TSC model, the distance mode profile-multi-branch dilated convolution network (DMP-MDNet). DMP-MDNet comprises two key modules aimed at bypassing complicated feature engineering and enhancing generalization capability. The first module, DMP, uses similarity measurement to encode scale- and magnitude-invariant temporal properties. Subsequently, the MDNet, equipped with multi-receptive field sizes, effectively leverages multi-granularity data for accurate classification. The effectiveness of our method is demonstrated through the analysis of a real-world body-in-white production dataset and various widely used public TSC datasets, showing promising applicability in actual manufacturing processes.

Keywords: Process monitoring; Anomaly recognition; Time series representation; Time series classification

1. Introduction

The emergence of cyber manufacturing in the manufacturing industries is attributed to rapid advancements in sensing, communication, and computation technologies. Massive amounts of data generated during the manufacturing process can now be sensed and recorded chronologically. Beyond advanced data acquisition, the key objective of cyber manufacturing underscores the efficient and in-depth utilization of these collected data, given their rich information about production status and quality [1]. Effective analysis of such data aids in evaluating process performance, enabling informed decision-making to ensure production quality [2]. The control chart, a commonly used statistical process control (SPC) tool in traditional production, can detect out-of-control data points by setting control limits (CL) and run rules for process parameters. However, the insufficiency of traditional SPC-based monitoring approaches in cyber manufacturing has been noted. Not all observations within the limits necessarily indicate normal status. Patterns over time, even within these limits, can signify unfavorable production behaviors, such as tool wear. The recognition of abnormal patterns and their underlying causes often requires manual analysis in SPC-based approaches. Furthermore, cyber manufacturing deals with production data exhibiting complex properties [3], displaying long-term, serially correlated, noisy, and dynamic characteristics. Conventional control charts, operating under the assumption of independent and identically distributed (i.i.d.) data, prove less effective in leveraging these production data.

Abnormal pattern recognition is a crucial aspect of manufacturing process monitoring. It involves identifying unexpected behaviors in process data that deviate from normal conditions. Identifying these unexpected data patterns is key to revealing abnormal events or states in the production process [4]. Algorithms such as support vector machines (SVM) and artificial neural networks (ANN) have been effectively employed to explore the mapping relationships between batch data features and their pattern labels [5,6]. However, the data in a cyber manufacturing environment typically comprises time series with dynamic and stochastic characteristics [7]. These “feature-to-label” methods often overlook the temporal aspect of process data. Two critical challenges are identified in the current paradigm:

1) The extraction of static and discrete features is based on the assumption of independent data, which fails to capture serial correlations in process data [8].

2) The manufacturing process might be disturbed by multiple anomalies simultaneously, resulting in multi-scale local patterns within the long-term time series generated during production. Most studies, however, classify fixed-length data from a single process into one specific label according to the global pattern, leading to the information loss of local patterns [9].

Therefore, time series classification (TSC) methods have been employed to overcome the limitations of existing abnormal pattern recognition methods [10]. Various TSC methods based on machine learning (ML) and deep learning (DL) have emerged. Typically, the efficacy of classification largely relies on the quality of the extracted temporal features. However, feature extraction from production process data in cyber manufacturing poses challenges. Gathering domain knowledge, crucial for high-quality temporal feature extraction, proves difficult in real-world manufacturing scenarios [11]. Additionally, custom-tailored features are often task-specific, which can hinder the generalization capability of classifiers across different production conditions. This is contrary to the goal of robust and resilient data analysis, a key requirement in cyber manufacturing.

In this study, we propose an abnormal pattern recognition method for manufacturing process

monitoring based on multi-scale TSC. The method involves slicing long-term time-series data from the complete production process using dynamically sized observation windows, and classifying them at multiple scales with our newly developed distance mode profile-multi-branch dilated convolution network (DMP-MDNet). The classification results from different observation scales are integrated into a probability diagram, providing clearer insights into the distribution of anomalies throughout the manufacturing process. The main contributions of this work are:

1) An abnormal pattern recognition method for manufacturing process monitoring utilizing multi-scale TSC is proposed. The recognition approach effectively transforms long-term process data with stochastic lengths into disentangled identification of abnormal manufacturing patterns across various scales. This helps in intuitively understanding the diverse anomalous conditions that arise during production.

2) By extracting robust scale-invariant and magnitude-invariant data information, a novel time series representation module DMP is first developed in the proposed time-series classifier DMP-MDNet to encode both local and global temporal trends in an interpretable form.

3) A classification module, MDNet, is designed to capture multi-granular data information. Featuring multiple receptive field sizes and dilated convolution, MDNet exhibits enhanced classification accuracy over models with a single receptive field.

4) The effectiveness and efficiency of the proposed method are validated through experiments conducted on a real-world body-in-white (BIW) production case and 25 public time series datasets.

The remainder of this paper is organized as follows: Section 2 reviews related work. Section 3 presents our abnormal manufacturing pattern recognition method based on multi-scale TSC. In Section 4, a BIW case study demonstrates the application of the proposed abnormal pattern recognition method. Section 5 details extensive experiments conducted on public datasets and discusses the results. Finally, Section 6 concludes the paper.

2. Literature review

2.1 Manufacturing process monitoring methods

2.1.1 Conventional manufacturing process monitoring methods

The control chart, a fundamental tool in SPC widely used by quality engineers in real-world production, graphically represents key parameters of the manufacturing process over time or production quantity. Classical charts include the Shewhart [12], cumulative sum (CUSUM) [13], and exponentially weighted moving average (EWMA) charts [14]. The Shewhart chart identifies an observation as out-of-control if it falls outside 3σ limits. However, this method, focusing only on the most recent observations, is mainly suitable for detecting large shifts. To increase sensitivity to smaller drifts, memory-based CUSUM and EWMA charts leverage historical data: CUSUM charts accumulate the sum of deviations between observations and the theoretical target, while EWMA charts monitor the moving average. Regardless of the chart type, CL must be set appropriately. Implementing control charts involves two phases. In Phase I, historical in-control data are analyzed to estimate the most suitable CL value. Subsequently, Phase II involves online monitoring based on the parameters established in Phase I. Typically, CL is determined through statistical analysis under the assumption of i.i.d data in traditional

approaches. However, this assumption becomes increasingly challenging to sustain as modern production processes grow in complexity.

Therefore, recent research has shifted towards leveraging artificial intelligence (AI) algorithms, such as KNN [15], SVM [16], and ANN [2,17], owing to their robust feature extraction capabilities. These algorithms can effectively utilize latent information in complex production data for accurate CL estimation. For instance, Yu et al. [17] proposed a process control method named the dynamic convolutional gated recurrent unit attention auto-encoder (DCGRUA-AE). This method uses spatiotemporal features and reconstruction losses extracted by DCGRUA-AE to calculate CL based on the T-square distribution and squared prediction error. AI algorithms contribute to enhanced monitoring in complex scenarios by extracting essential information from massive, dynamic, and serially dependent data.

However, both traditional and AI-enhanced control charts primarily focus on whether data points fall within set limits. Note that single abnormal points may result from measurement errors or noise, and not all points falling within CL necessarily indicate a normal state. The control chart alone is inadequate for addressing more intricate abnormal patterns [18]. Abnormal patterns not only signify product defects, but also are closely related to the status of the manufacturing devices and thus the maintenance of the production line [19]. The lack of effective pattern analytical tools generally necessitates the participation of skilled labor to manually identify the quality problem and its corresponding cause. This runs counter to the goal of efficient, resilient, sustainable maintenance in Industry 4.0. The time-consuming analysis process may require more experience and technical support with excessive costs when dealing with complex industrial processes. Furthermore, the inherently subjective nature of manual analysis can lead to delayed detection of production faults in their early stages, escalating into reduced productivity, undesirable breakdown, unplanned downtime, and other issues that jeopardize the economic sustainability of the production line [20]. Especially in the automotive manufacturing industry from which our case study comes, the pursuit of higher product quality at lower manufacturing costs is paramount [21] given the capital-intensive nature of the industry and high non-renewable energy consumption during production [22,23]. While popular quality assurance techniques such as Six Sigma and lean methods [24,25], as well as AI-based anomaly detection approaches [26,27], are capable of detecting anomalies to trigger maintenance, they remain insufficient to recognize the maintenance need based on identifying abnormal cases and causes promptly. Thus, more advanced process monitoring approaches that can identify specific behaviors presented by the continuous collection of data points are still needed to support proactive and informed maintenance operations.

2.1.2 Pattern classification methods for manufacturing process monitoring

The limitations of conventional monitoring approaches have led to the investigation of AI-based classification methods for pattern recognition. As a key enabler for big data analytics, intelligent algorithms prove to be feasible for extracting data patterns that foresee future production faults or device degradation and, therefore, keep the production lines well-maintained and resource-efficient [28,29]. Lee et al. [30] utilized a deep belief network to establish the correlation between defect labels and process data features, such as torque signals, enabling fault detection in the manufacturing of vehicle high-intensity discharge (HID) headlights and cable modules. To manage the online inspection data of BIW with high volume and low signal-to-noise ratio, Liu et al. [31] first denoised the data, followed by processing the shape and statistical features through a backpropagation network for pattern recognition.

However, the statistical features commonly used in these methods have i.i.d. data premise, an assumption hardly tenable to temporally correlated production processes.

More research has focused on investigating effective approaches to handle the temporal nature of the production process data. Some works try to make full use of valuable temporal information by enhancing feature extraction procedures. Liao et al. [32] first transformed the acoustic emission signals from the cutting process into scalograms, followed by conducting pattern classification using various vision backbones, such as VGG-16. Mahato et al. [33] adopted the dynamic time warping (DTW) approach to extract temporal features from temperature signals of the melt pool and then applied a KNN classifier to monitor the additive manufacturing process. Qian et al. [34] designed a sequence feature extraction network. Through denoise autoencoder and DTW-based local feature extraction, discriminative temporal features were constructed for fault pattern recognition in the Tennessee process benchmark.

Utilizing special network structures such as 1D convolutional neural network (CNN) and recurrent neural network (RNN), other studies attempt to empower the classification models to directly handle sequential inputs. Liu et al. [8] successfully recognized variation patterns in BIW measurement data using a long short-term memory network, demonstrating efficient utilization of temporal information. Yousef et al. [35] proposed an end-to-end process monitoring approach based on visual analytics. Temporal patterns were visually classified by 2D convolution operations after the shallow temporal features extracted by 1D CNN were reshaped into 2D images. The proposed monitoring approach was successfully applied to identify faults in the pyrometallurgical smelting process. Zhao et al. [36] combined adaptive multi-scale CNN, LSTM, and channel attention mechanisms to automatically mine contextual information, local features, and global temporal dependencies from industrial process data, enabling fault diagnosis for the Tennessee process.

The importance of temporal characteristics in manufacturing process data is substantial. However, there is a notable scarcity of studies applying TSC methods to process monitoring. Moreover, most classification-based monitoring studies typically assign a single label to one sequence, overlooking the fact that long-term processes may exhibit different patterns at multiple observation scales.

2.2 Time series classification methods

TSC methods can be categorized into three main types based on the data information utilized as classification criteria, including distance-based, hand-crafted feature-based, and deep feature-based.

Distance (or similarity)-based methods make classifications based on the distance between samples. Typical distance metrics include Euclidean distance [37], DTW [38,39], edit distance on real penalty [40], edit distance on real sequence [41], etc. Although generally easy to implement, distance-based methods can be quite sensitive to the noise inherent in real production.

Hand-crafted feature-based methods have been developed to mitigate the impact of redundant information in TSC. Various feature extraction techniques have been proposed to capture the essence of time series more accurately. Shapelet-based methods [42,43] focus on retaining subsequences that demonstrate the highest discriminative capability. The classification performance is improved by concentrating on the most significant parts of the time series. However, owing to the computational complexity of discovering shapelets, these methods are more suitable for shorter time series. For long-

term TSC tasks, symbolic or dictionary-based features, such as Symbolic Fourier Approximation and Symbolic Aggregate Approximation are preferred [44,45]. In these methods, time series are segmented and encoded using a certain dictionary. The sequence is converted into a shorter string and classified based on word frequency. Feature engineering is another approach to construct hand-crafted features. This includes simple statistical features such as mean and variance [46], as well as more complex ones derived from wavelet transform, spectral analysis, and phase space reconstruction [32,47]. Inspired by advancements in computer vision, converting time series into 2D feature maps, such as Poincaré and recurrence plots [48,49], has also gained prominence. In addition, some studies focus on extracting features from specific intervals of the series rather than the entire series. The features from these intervals are then combined into feature vectors for classification [50,51]. However, the effectiveness of classifiers heavily depends on the quality of the selected features. Manually designing time series features is laborious and requires specialized domain knowledge. Designing effective features that retain valuable information while discarding the superfluous remains a significant challenge [52].

End-to-end methods rely on deep networks for automatic feature extraction. CNN [53,54], RNN [55,56], attention models [57,58], and their variants such as temporal convolution network [59], are widely used in TSC problems. Moreover, multimodal models that combine various network types are also a significant research focus [60,61]. Similar to ensemble learning which combines the results of multiple classifiers [51,62], multimodal networks integrate the diverse learning capabilities of different models to process data from multiple perspectives. While end-to-end TSC methods have shown great promise, they do have certain limitations. Firstly, the automatically extracted features are not always beneficial for the final classification, as shallow features may contain noise or other irrelevant information. Secondly, the interpretability of these features is often insufficient [63].

In addition, the dynamic data conditions inherent in cyber manufacturing impose more stringent demands on the generalizability of data analytical tools. However, many TSC methods are bound by specific training constraints. Consequently, when the data distribution shifts as a result of alterations in production conditions, classifiers pre-trained for specific processes are prone to failure. Therefore, addressing the challenge of extracting robust latent information from collected time series data is essential when applying TSC methods to abnormal manufacturing pattern recognition.

3. Abnormal pattern recognition based on multi-scale TSC for manufacturing process monitoring

3.1 Overall framework of multi-scale abnormal pattern recognition method

According to classical control chart theory [64], temporal data in the manufacturing process typically display six patterns, as shown in **Fig. 1**: normal patterns (NOP), cyclic patterns (CYC), increasing trend patterns (ITP), decreasing trend patterns (DTP), upward shift patterns (USP), and downward shift patterns (DSP). These patterns indicate changes in production status due to different causes. Accurately recognizing these fluctuation patterns is crucial for effectively identifying the root causes.

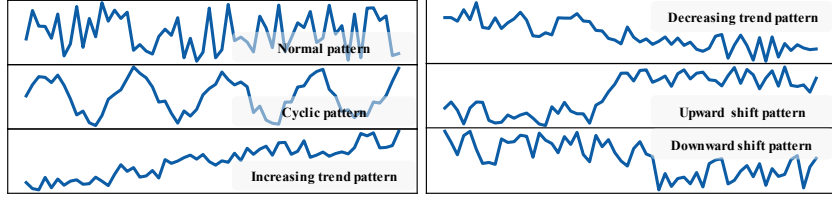


Fig. 1 Common patterns in manufacturing process.

Numerous control chart pattern recognition methods overlook local patterns in long-term processes. For instance, as illustrated in **Fig. 2**, the geometric feature inspection data of 800 consecutively produced BIWs displays periodicity when viewed in a large-scale window. However, a downward shift becomes apparent when the observation window is reduced to 200. Further narrowing the observation scale to 50 reveals a decreasing trend. This indicates that excessively large observation windows may fail to identify instantaneous changes, while smaller windows might overlook long-term trend changes.

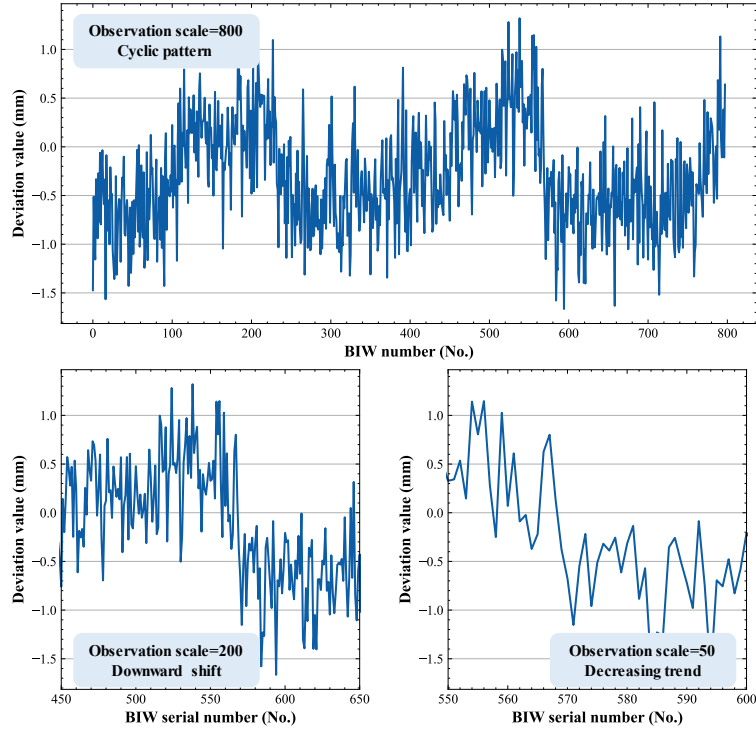


Fig. 2 Patterns under different observation scales.

Motivated by this challenge, this study introduces an abnormal manufacturing pattern recognition method based on multi-scale TSC. The overall framework is presented in **Fig. 3** and **Algorithm 1**. The temporal signals of key quality features captured during the production process are initially sliced. Once the start point for observation is specified, observation windows with expanding scales are used for data slicing in order to analyze the dynamic property of the production process at various time scales, for instance, by minute, by hour, or by day. Then, the data slices are separately inputted into a TSC model, DMP-MDNet. In this model, each time-series slice is first encoded as a 2D temporal representation and then mapped into a 1×6 vector by the classification module with multiple receptive sizes fused. Each element of the output vector represents the categorical distribution of the input slice belonging to different temporal pattern categories. Finally, the categorical distribution vectors corresponding to all time-series slices are stacked in the order of observation scale and compiled into an $n \times 6$ anomaly distribution diagram. This diagram effectively illustrates the comprehensive changes in manufacturing status along

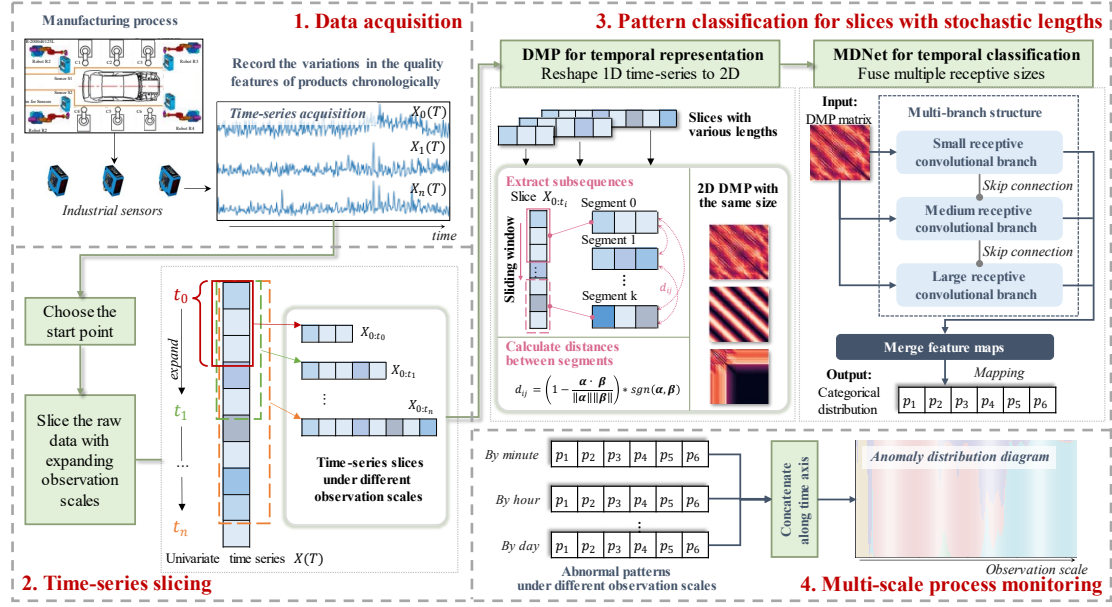


Fig. 3 Overall framework of the proposed multi-scale abnormal pattern recognition method.

Algorithm 1 Multi-scale TSC-based abnormal pattern recognition for manufacturing process monitoring

Input:

$X(T) = (x_0, x_1, x_2, \dots, x_T)$: Univariate time series of key quality features collected during production

t_0 : Basic observation scale

Δt : Expanding stride of the observation scales

L : Size of DMP matrix

s : Sliding window step

1: Slice the input univariate time series by dynamically sized observation scales:

1.1: for $i = 0$ to $\left\lfloor \frac{T-t_0}{\Delta t} \right\rfloor$ do

1.2: slice $_i = X_{0:t_i} = (x_0, x_1, x_2, \dots, x_{t_0+i})$

1.3: end for

1.4: Obtain all time-series slices $\mathcal{S} = \{\text{slice}_0, \text{slice}_1, \text{slice}_2, \dots\}$

2: Perform pattern recognition for each time-series slice with DMP-MDNet model:

2.1: for $i = 0$ to $\left\lfloor \frac{T-t_0}{\Delta t} \right\rfloor$ do

2.2: Calculate the sliding window size: $w = t_i - s(L-1)$

2.3: for $j = 0$ to $\left\lfloor \frac{t_i-w}{s} + 1 \right\rfloor$ do

2.4: segment $_j^{(i)} = (x_{js}, x_{1+js}, x_{2+js}, \dots, x_{w+js-1})$

2.5: end for

2.6: Obtain all subsequences in the time-series slice $X_{0:t_i}$

2.7: DMP transformation for $X_{0:t_i}$ by calculating the similarities between subsequences

$$D_i = [d_{ml}^{(i)}] = \left(1 - \frac{\text{segment}_m^{(i)} \cdot \text{segment}_l^{(i)}}{\|\text{segment}_m^{(i)}\| \|\text{segment}_l^{(i)}\|} \right) * \text{sgn}(\text{segment}_m^{(i)}, \text{segment}_l^{(i)})$$

2.8: Conduct classification with DMP matrix as the input of MDNet: $o_i = \text{MDNet}(D_i)$

2.9: end for

3: Concatenate the categorical distribution vectors corresponding to all time-series slices

Output:

$P = [o_1, o_2, \dots, o_{\left\lfloor \frac{T-t_0}{\Delta t} \right\rfloor}]$: Pattern distribution diagram

3.2 Time series classifier based on DMP-MDNet

A TSC model, DMP-MDNet, requiring minimal domain knowledge, is proposed for the pattern classification of time-series slices. It comprises two primary modules: the DMP module for temporal representation transformation and MDNet for classification.

3.2.1 DMP for temporal representation

As illustrated in **Fig. 1**, the fundamental difference between typical data patterns in the production process lies in their change trends, rather than their absolute values. To capture this, a time-series representation module named DMP is proposed. It interpretably encodes scale- and magnitude-invariant properties by measuring the distances between all subsequences. The calculation process is as follows:

First, the time-series slice with length N , denoted as $X_{0:N-1} = (x_0, x_1, x_2, \dots, x_{N-1})$, is segmented into a set of subsequences $\mathcal{S} = \{s_{i,w}\}$ using a sliding window. The sliding window, with a width w , starts at the initial point of $X_{0:N-1}$ and extracts the first subsequence $s_{0,w} = (x_0, x_1, x_2, \dots, x_{w-1})$. The window then moves forward by a step s to extract the second subsequence $s_{1,w} = (x_s, x_{1+s}, x_{2+s}, \dots, x_{w+s-1})$. This process is repeated until the entire time-series slice is covered. For a slice of length N , n subsequences can be extracted in this manner. The window sizes and sliding steps, as outlined in Eq. (1), are automatically adjusted for data slices of different lengths to maintain consistent sizes of the DMP matrices.

$$n = \left\lfloor \frac{N - w}{s} + 1 \right\rfloor \quad (1)$$

After extracting the subsequences, their similarities are measured to encode the DMP map $D = [d_{ij}]$. d_{ij} represents the distance between the i^{th} subsequence α and j^{th} subsequence β . Considering the fundamental differences in common patterns of industrial process data, the similarity metric used in DMP calculation should meet two criteria: (1) it should effectively express tendency information, and (2) it should be insensitive to the absolute value and noise level of the time series. The formula demonstrates that the cosine, Canberra, Bray-Curtis, and correlation distances are all insensitive to absolute values. However, both the Canberra and Bray-Curtis distances are susceptible to noise because they calculate pointwise distances. The correlation distance is also not suitable, as it fails to effectively distinguish between NOP, ITP, and DTP patterns, given that the subsequences in these three patterns are linearly correlated.

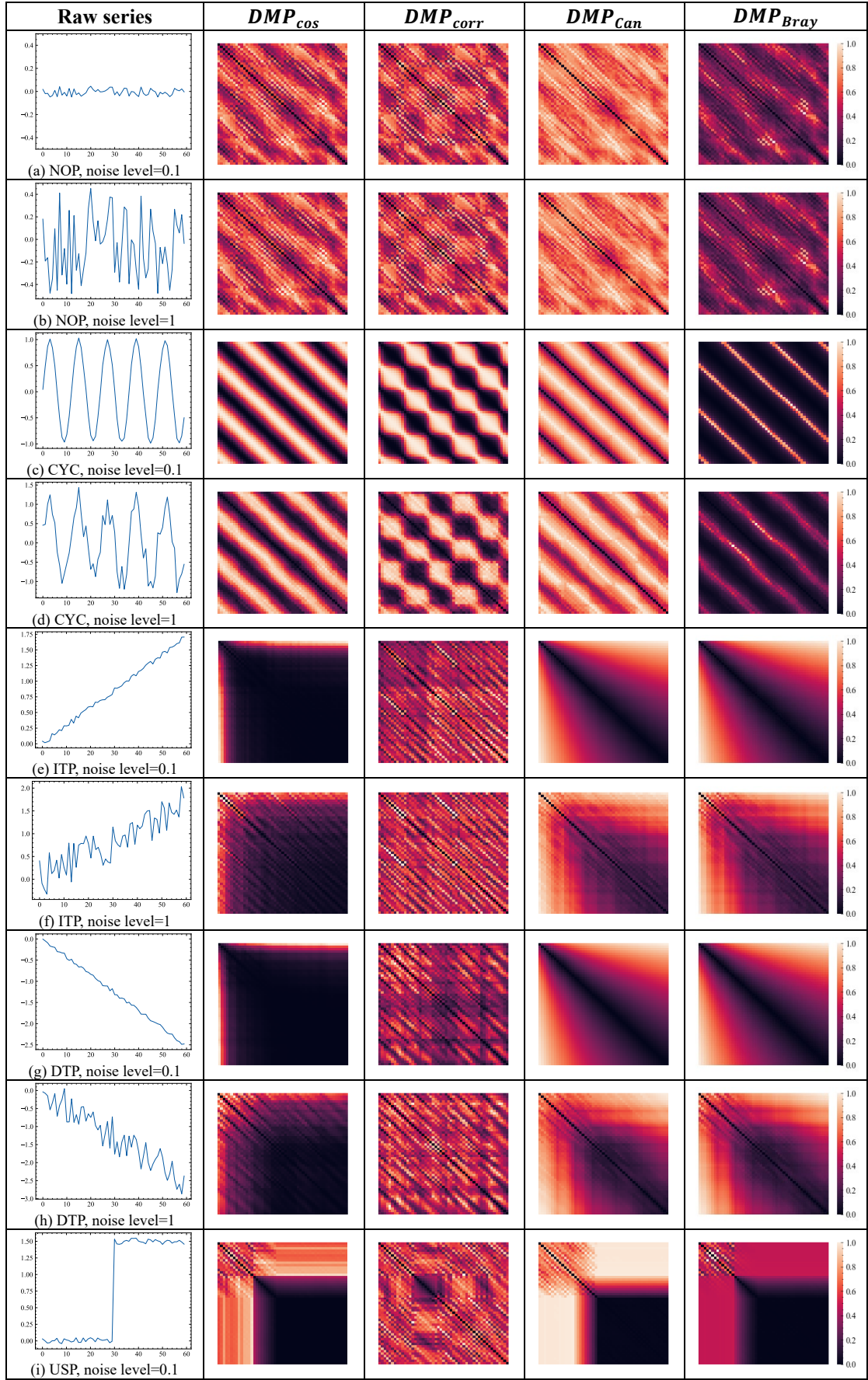
$$\text{Cosine distance: } d_{ij} = 1 - \frac{\alpha \beta}{\|\alpha\| \|\beta\|} \quad (2)$$

$$\text{Canberra distance: } d_{ij} = \sum_{k=1}^n \frac{|x_k - y_k|}{|x_k| + |y_k|}, x_k \in \alpha, y_k \in \beta \quad (3)$$

$$\text{Bray-Curtis distance: } d_{ij} = \frac{\sum_{k=1}^n |x_k - y_k|}{\sum_{k=1}^n x_k + \sum_{k=1}^n y_k}, x_k \in \alpha, y_k \in \beta \quad (4)$$

$$\text{Correlation distance: } d_{ij} = 1 - \frac{\text{cov}(\alpha, \beta)}{\sigma(\alpha) \cdot \sigma(\beta)} \quad (5)$$

Fig. 4 displays the normalized heat maps of DMP matrices using different similarity metrics. To assess robustness, white noises $\varepsilon_i \sim N(\mu, \sigma^2)$ at different levels k were added to the raw series. The darker the color in the heat map, the more similar the two subsequences are. The visualization results align with those of the previous analysis. Among the four metrics evaluated, the cosine distance emerges as the most appropriate for DMP calculations. The cosine distance-based DMP matrices of different patterns exhibit distinguishable characteristics, clearly and intuitively reflecting the global and local trends of the time series. Furthermore, the DMP maps remain largely unchanged with increasing noise levels.



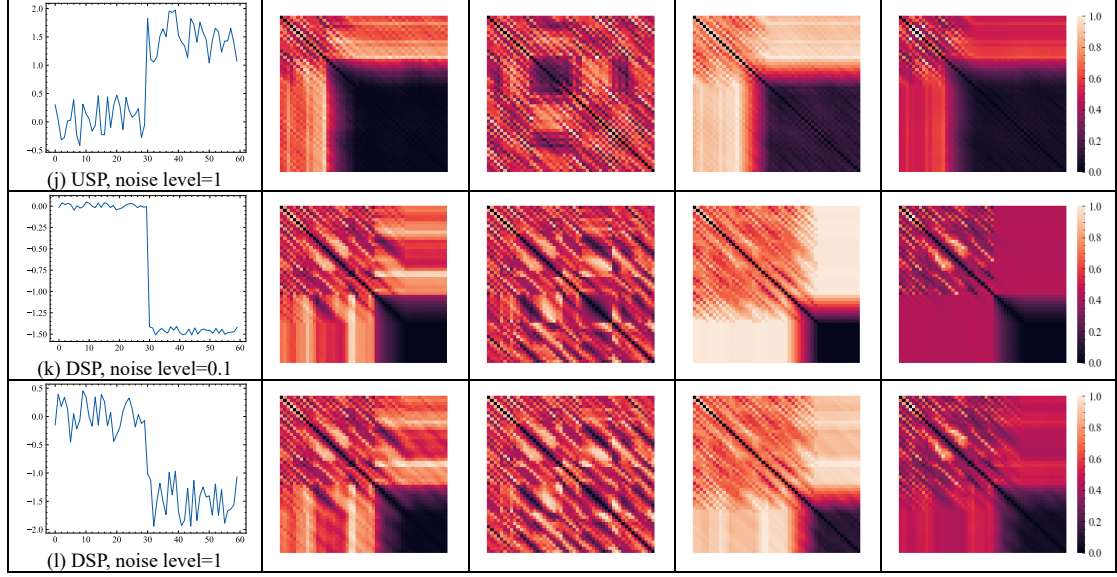


Fig. 4 DMP heatmaps with various similarity metrics and different noise levels.

While the cosine distance effectively calculates the relative distance between subsequences, it does not provide information about the precedence order of subsequences α and β in the entire time-series slice. As a result, distinguishing between ITP and DTP, or USP and DSP, becomes challenging. To improve the discrimination between different patterns, this paper introduces a temporally aware cosine distance for DMP calculation, as shown in Eq. (6). $sgn(\alpha, \beta)$ denotes a sign function designed to distinguish the order of subsequences in the overall data slice. $ones = \{1, 1, 1, \dots, 1\}$ in Eq. (7) is a reference vector in the w -dimensional space where α and β are located.

$$d_{ij} = \left(1 - \frac{\alpha \cdot \beta}{\|\alpha\| \|\beta\|}\right) * sgn(\alpha, \beta) \quad (6)$$

$$gn(\alpha, \beta) = \frac{dist_{cos}(\alpha, ones) - dist_{cos}(\beta, ones)}{|dist_{cos}(\alpha, ones) - dist_{cos}(\beta, ones)|} \quad (7)$$

As demonstrated in **Fig. 5**, differentiating the DMP maps of patterns with opposite trends using the original cosine distance is problematic. However, with the temporally aware cosine distance-based DMP maps, distinctions become clear, as the DMP maps are flipped along the diagonal, indicating the temporal order of the patterns.

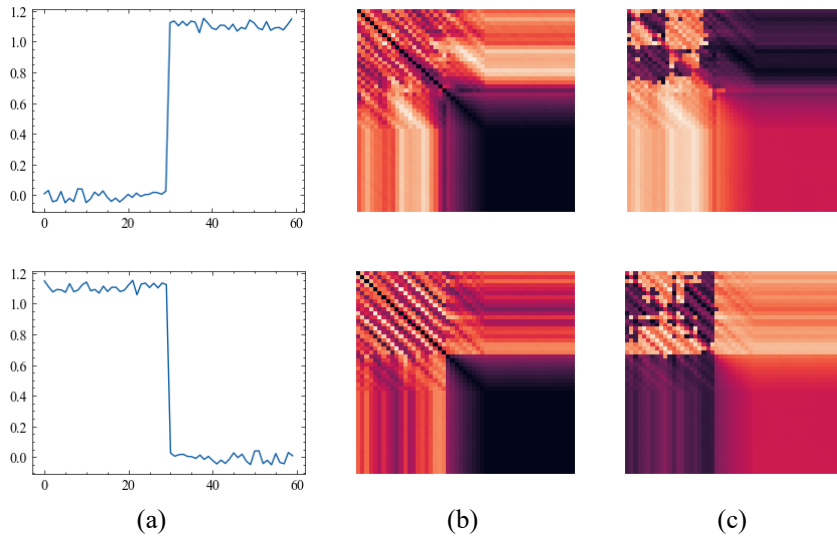


Fig. 5 DMP maps of two opposite patterns. (a) Raw series with opposite patterns; (b) DMP maps with

conventional cosine distance; (c) DMP maps with temporal-aware cosine distance.

3.2.2 MDNet for classifying DMP matrices

The DMP calculations transform the original 1D time-series slice into a 2D feature map. Drawing inspiration from the success of 2D CNNs in image classification, this paper introduces a network with a multi-branch structure and dilated convolution mechanism to analyze high-level information and ultimately provide classification results.

The size of the receptive field, determined by the kernel size, is important for the performance of CNN-based models. Traditional CNNs, however, typically employ a fixed kernel size, leading to a single and limited receptive field that lacks the capability to fuse multi-scale information. To address this, several studies have proposed CNN structures with multi-receptive fields [65,66]. Building on this concept, this paper has designed a classifier with small, medium, and large receptive field branches. Feature maps in the middle layers are aggregated across branches to enable multi-level information fusion without increasing the number of network layers. In addition, skip connections are added within each branch to facilitate effective network training, even with small datasets. The structure of this network is illustrated in Fig. 6.

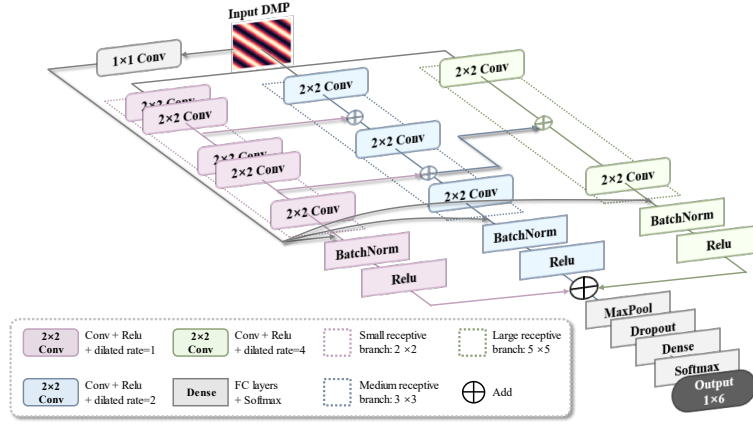


Fig. 6 MDNet structure.

To enhance computational efficiency, this paper employs dilated convolution. Dilated convolution introduces intervals within the normal kernel, enlarging the receptive field without significantly increasing the number of parameters. An ordinary convolution of size $k \times k$ with a dilation rate d has an effective receptive field size of:

$$k' = k + (k - 1)(d - 1) \quad (8)$$

The output size of one convolutional layer can be determined using Eq. (9). With fewer dilated layers, the same effect can be achieved as stacking multiple conventional convolution layers. Therefore, dilated convolution enables a rapid increase in the receptive field for more effective information exploitation, while maintaining computational efficiency.

$$output_size = 1 + \frac{input_size + 2padding - kernel_size}{stride} \quad (9)$$

The DMP matrix D serves as the input for the MDNet, with the 1×6 categorical probability vector as the output $o(D)$. Each dimension in the array corresponds to a pattern category, and its value represents the probability of the sample belonging to that category.

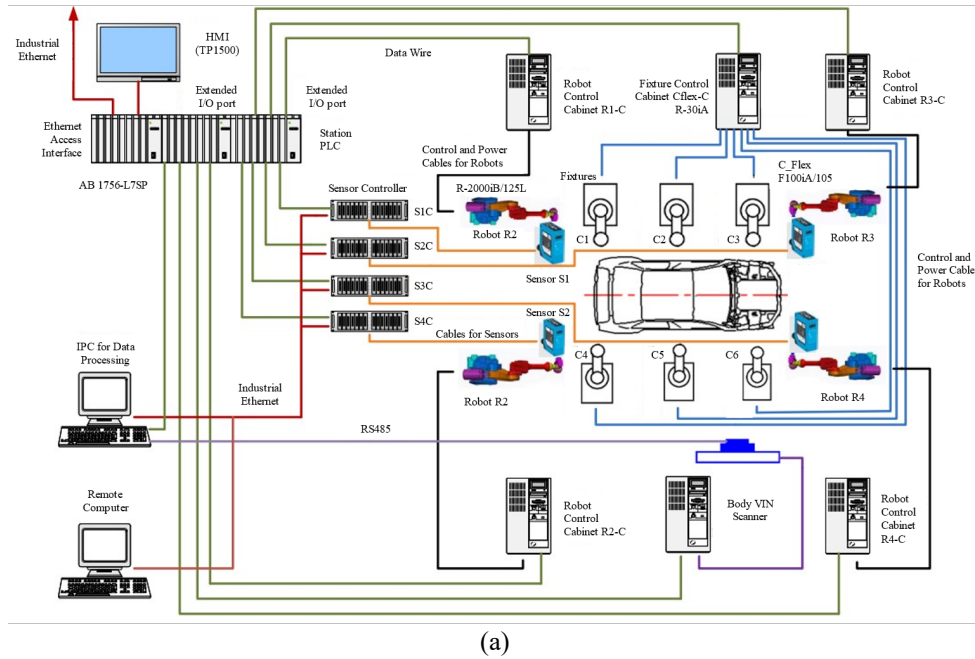
$$\begin{aligned}
o(D) &= MDNet(D) \\
&= f_{MDNet} \left(Conv_{small}(D), Conv_{middle}(D), Conv_{large}(D) \right) \\
&= \{P_{NOP}, P_{CYC}, P_{ITP}, P_{DTP}, P_{USP}, P_{DSP}\}
\end{aligned} \tag{10}$$

4. Case study

This section details an actual BIW production case. The implementation of the proposed abnormal pattern recognition method was conducted on a computer equipped with a Microsoft Windows 11 operating system, an Intel-i7-11700F CPU, 16GB RAM, and a GPU of NVIDIA GeForce RTX 3060Ti. The software framework was TensorFlow 2.10.0, operating on a Python 3.8 backend.

4.1 Data description

Online inspection data from SAIC-GM-Wuling automobile were utilized to evaluate the feasibility of the proposed multi-scale abnormal pattern recognition method. This data was gathered using a digital twin-based BIW online inspection system. The experimental setup, as shown in Fig. 7, involves the following steps: After completing the welding process, the BIW is transported to the inspection site and positioned using the FANUC C-flex fixture system. Four FANUC GLOBAL3 robots equipped with laser scanners then collect the cloud-point data of the BIW. Subsequently, dimensions of critical features such as holes, grooves, and folded edges on the front windshield are extracted, and the geometric deviations from theoretical values are calculated and output. Additional details on the system setup and experimental environment are available in a prior study [67].



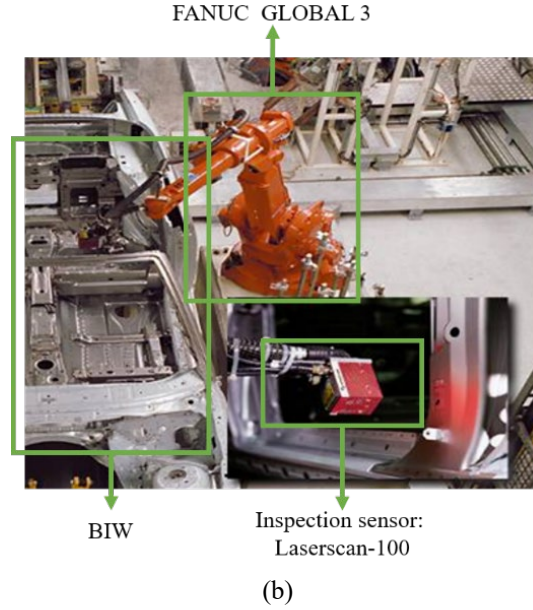


Fig. 7 Hardware module of online inspection digital twin system. (a) Experimental setup; (b) Inspection site setup.

Process quality evaluation can be effectively conducted through a detailed analysis of these geometric deviation sequences. For this study, a total of 1000 sequence samples were selected from the collected data to create a pattern dataset. This dataset encompasses the six fluctuation patterns discussed in Section 3.1, including 192 NOP, 85 CYC, 170 ITP, 169 DTP, 176 USP, and 208 DSP samples. Note that the lengths of these samples vary, with the longest sequence being 798 units and the shortest being just 50 units. Additionally, real-world data often deviate from ideal conditions. For instance, while the ideal CYC pattern is strictly sinusoidal, such patterns are rare in actual production settings, where the signal usually exhibits quasi-periodicity. The observed waveforms in production are more often square-shaped or ramp-like in nature.

4.2 DMP-MDNet training

The time-series classifier DMP-MDNet was initially trained using BIW data. It utilized the 1D time series as input, which was then transformed into a 2D DMP matrix and subsequently classified using MDNet. The training parameters were configured as follows: the dataset split ratio was set at 9:1, DMP_size was 30, the learning rate was 0.01 with a 0.5 decay factor and 3 patience, batch_size was 64, and the training epoch was 64. The classification performance was evaluated using metrics such as *Accuracy*, *Precision*, *Recall* and *F1 – Score*. As depicted in **Fig. 8(a)** and **Fig. 8(b)**, the model demonstrated convergence after approximately 60 training epochs. The final accuracy on the test set reached 93.04%. **Fig. 8(c)** presents the confusion matrix, illustrating the number of samples correctly and incorrectly classified for the six patterns. For instance, in the first row, it can be observed that 92.9% of the NOP samples were accurately classified, while 7.1% were misclassified as DTP. Notably, the accuracy for all six patterns exceeded 90%. Despite the presence of real data distribution imperfections and noise, DMP-MDNet showed strong performance.

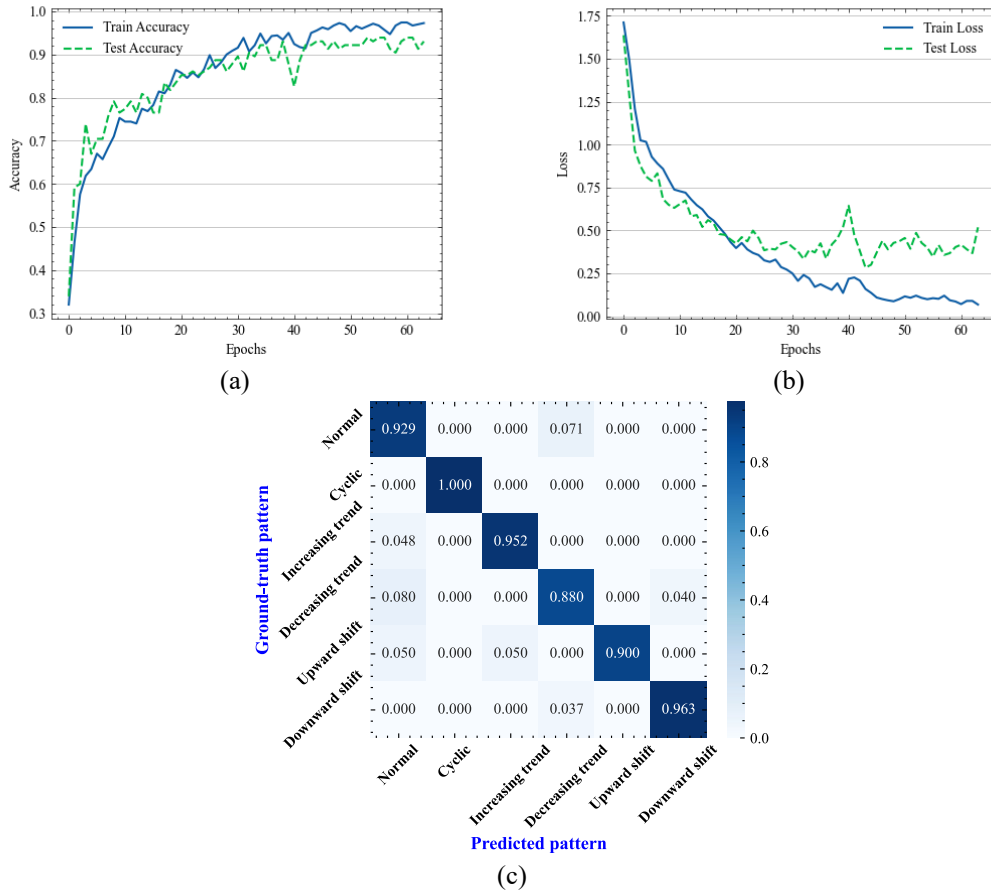


Fig. 8 DMP-MDNet performance on BIW production data. (a) Learning curve of accuracy; (b) Learning curve of loss; (c) Classification results.

4.3 Application in real manufacturing process

Fig. 9(a) illustrates the basic data analysis functions originally available in the online inspection system mentioned in Section 4.1. This original data analysis module can automatically trigger alarms for abnormal points of geometrical deviations exceeding ± 1.5 mm limits and calculate SPC statistics. However, these rudimentary functions cannot solely uncover concealed abnormal patterns in the data, owing to the absence of synergy between industrial big data and smart analytics. An example is presented in **Fig. 9**. **Fig. 9(c)** displays the geometric deviations of measuring point No. 10 on the front frame (depicted in **Fig. 9(b)**) for 100 sequentially manufactured BIWs. Despite all data points falling within the permissible tolerance range, the deviation curve manifests an abnormal shift, typically caused by sudden changes such as fixture loosening. Unfortunately, the original basic functions grounded in traditional statistical analysis consider this situation normal and do not issue an alarm. If these abnormal patterns are not promptly identified, the impact of abnormal factors gradually increases, resulting in a reduced rate of high-quality products. Therefore, to address the shortcomings of the basic data analysis functions, our work has been integrated into this online inspection system as a new function to identify these concealed patterns.

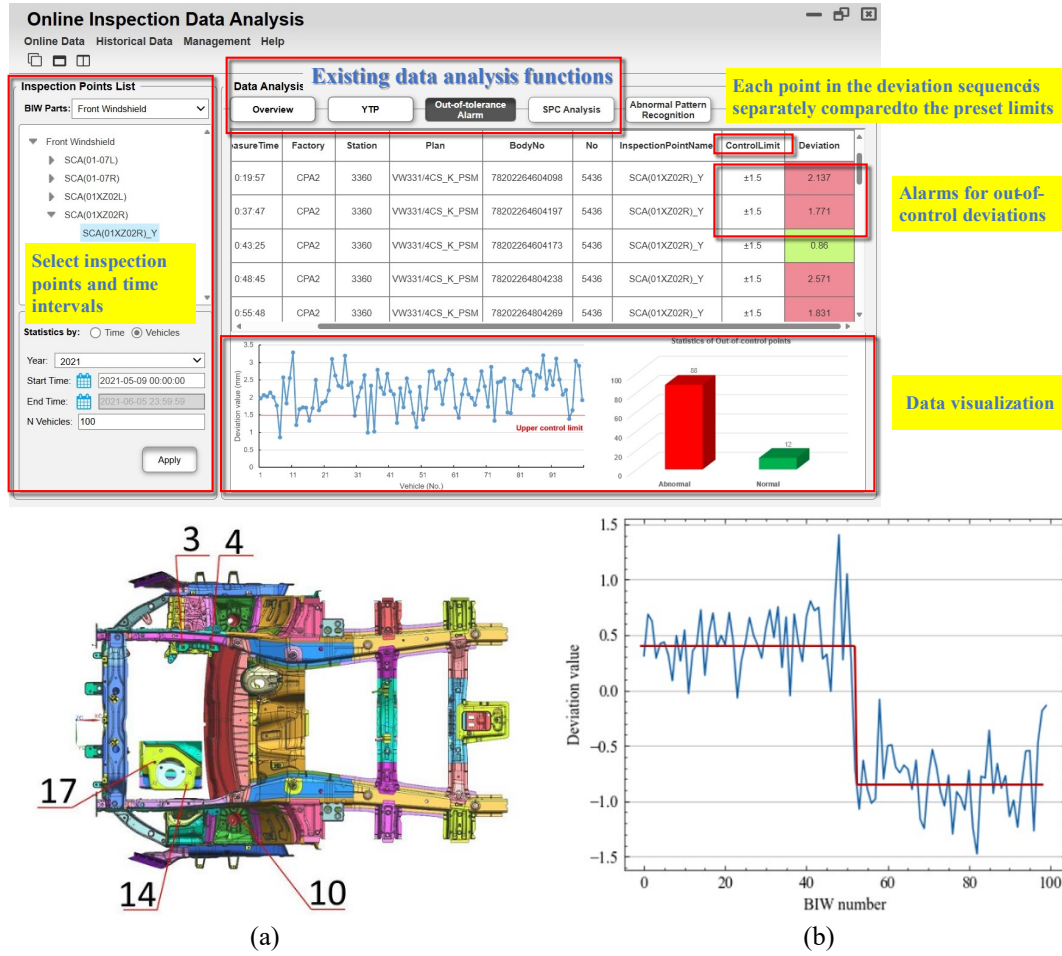


Fig. 9 Illustration on the limitation of the current online inspection system. (a) Basic data analysis functions of the system; (b) Inspection points layout; (c) Abnormal pattern visualization.

In this newly developed functional module, the pre-trained DMP-MDNet was employed for multi-scale abnormal pattern recognition. The deviation sequence from a specific time period underwent slicing using multi-scale observation scales. The minimum observation scale remained fixed at 30, while the maximum scale equated to the length of the entire series. The initial points of the observation scales were set as the first point of the sequence. Consequently, $(N - 29)$ slices of varying lengths were extracted from a sequence with a total length of N . These slices were input into the DMP-MDNet in the specified order, and pattern probabilities at different scales were computed. The ultimate results constituted an $(N - 29) \times 6$ matrix, which was subsequently visualized as shown in Fig. 10. The geometric deviation sequence on the side hinge of 300 BIWs consecutively produced is displayed in the middle of the interface. Using the proposed process monitoring method, the bottom graph depicts the corresponding distribution of anomalies that occurred throughout the production process of these 300 BIWs. In the anomaly distribution diagram, the x-axis represents the observation scale, while the y-axis denotes the proportion of the production status attributed to the six patterns at each instance. Normal fluctuations predominated when the observation window was small. As the window size gradually increased, the data started exhibiting an upward trend. Once the window width exceeded 150, the proportion of cyclic anomalies increased, becoming the predominant anomaly pattern. These monitoring outcomes aligned with the actual scenario. Visualizing the distribution diagram provides an intuitive understanding of all types of anomalies throughout the entire process. Below the anomaly distribution diagram, the percentages of local anomalies within the queried observation intervals can be found. For abnormal

patterns with high percentages, the system provides alarm information, including the commencement and conclusion times of the abnormal status, along with potential causes.

The results of this case study affirm the effectiveness of the proposed multi-scale abnormal pattern recognition method. This method can recognize multiple patterns of varying scales within a queried time series, offering insights into abnormal changes in production status and their underlying causes. These insights can be utilized for quality control feedback at the physical production site.

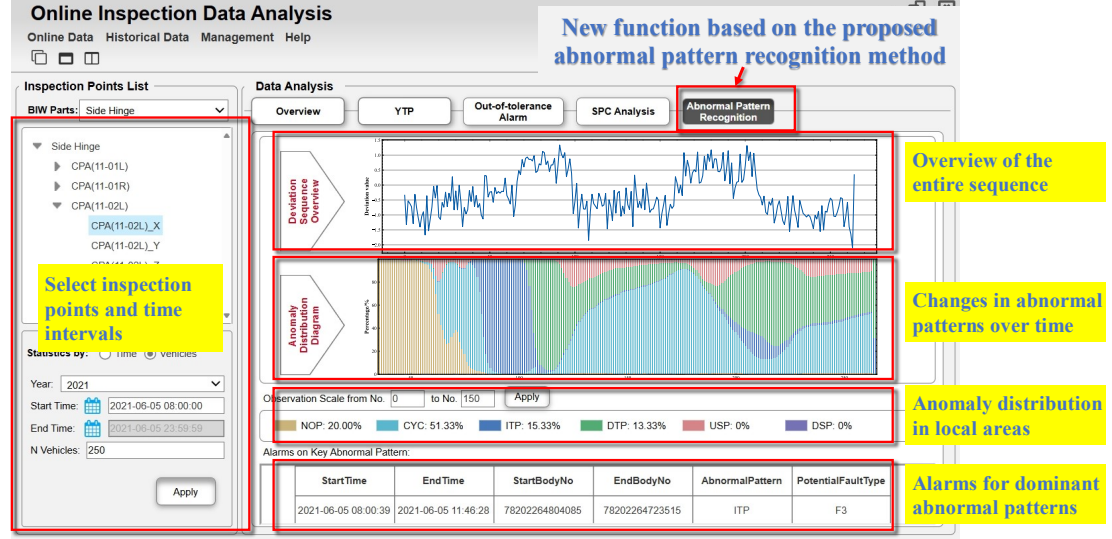


Fig. 10 Abnormal pattern recognition results

5. Comparison and discussion

In this section, a simulated control chart dataset was first employed to verify the need for multi-scale classification in the proposed pattern recognition method. Subsequently, comparative experiments were carried out using various public datasets sourced from the UCR archive [68]. These experiments aimed to delve deeper into the performance of the proposed TSC model DMP-MDNet, including its generalization capability, processing speed, and classification effectiveness.

5.1 Effect of dynamically sized observation window

This section delves into the effect of multi-scale observation windows on the proposed pattern recognition method. Monte Carlo simulation was utilized to generate control chart samples in accordance with [69]. The sequence length was fixed at 150 data points, with each sample containing only one specific pattern type.

Following the data simulation, each sample underwent classification under different observation scales, specifically set at 30, 60, 90, 120, and 150. For instance, with an observation scale of 30, a sliding window of $width = 30$ moved along the time axis from $t = 0$ to $t = 150$ with a step of $s = 1$, extracting time series segments. In the case of a sample $X(T)$ with a $length = 150$, 121 subsequences of length 30 were extracted. These extracted segments were then fed into the pre-trained DMP-MDNet. This process generated a total of 121 classification results, each representing the sectional patterns at different time points. The pattern label with the highest occurrence frequency was selected as the final classification result for $X(T)$ and compared with its true label. Similar procedures were repeated for other observation scales.

The classification performance of all patterns depicted in **Fig. 11** exhibited improvement as the observation scale increased. Notably, when the observation scale reached 150, that is, $X(T)$ was classified as a whole, the *Recall* metric reached its peak. Specifically, positive recognition results were achieved for NOP, USP, and DSP, even with small observation scales. Conversely, the classification effectiveness for CYC was notably lower with small windows. For instance, when $w = 30$, the *Recall* for CYC was only 0.385, but it increased significantly as the window size expanded. This implies that only a few samples with small cycles could be correctly identified with a small observation scale, which is inadequate for recognizing CYC samples occurring over an extended duration. These findings underscore the importance of dynamically sized observation windows, which enable a comprehensive understanding of multi-scale anomalies distributed throughout the manufacturing process. Small windows are adept at capturing instantaneous changes, whereas a broader observation scale is necessary for capturing long-term trends.

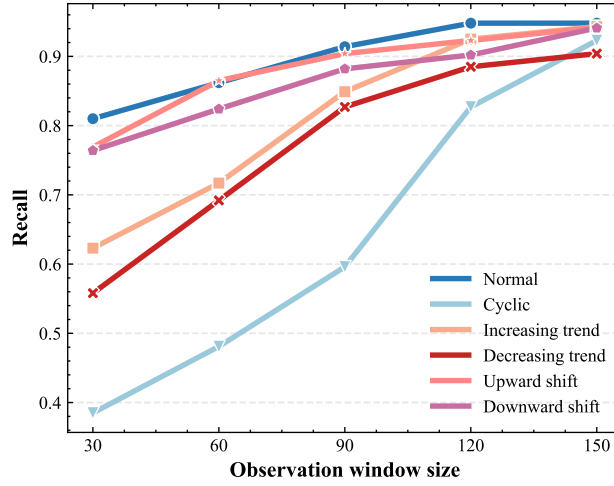


Fig. 11 Classification recall under observation windows with different sizes.

5.2 Generalization ability test

A public control chart dataset called “synthetic control chart time series” (SCCTS) is utilized in this section to test the generalization capability of DMP-MDNet. The SCCTS dataset contains 600 samples, each with a length of 60. In contrast to the BIW data, the SCCTS data exhibit varying magnitudes. While BIW samples fluctuate within the $(-3, +3)$ interval, the SCCTS data encompass a broader range of values. The maximum value in the SCCTS data can reach 63.82, while the minimum value can be as low as -5.11.

DMP-MDNet, trained using BIW production data, was directly applied to the SCCTS data without retraining, and no preprocessing steps such as normalization or denoising were performed. The final accuracy achieved was 90.21%, demonstrating strong generalizability of the DMP-MDNet. The detailed classification results are depicted in **Fig. 12**. The accuracy for NOP, ITP, DTP, USP, and DSP was approximately 95%. However, the accuracy for CYC was only 64%. One possible reason for this discrepancy is that the CYC samples in the SCCTS dataset are artificially generated and differ from the quasi-periodic CYC samples in the BIW data. Because the CYC samples used in the training and testing stages exhibit substantial differences, the model parameters obtained from the BIW data may not be entirely suitable for the SCCTS data.

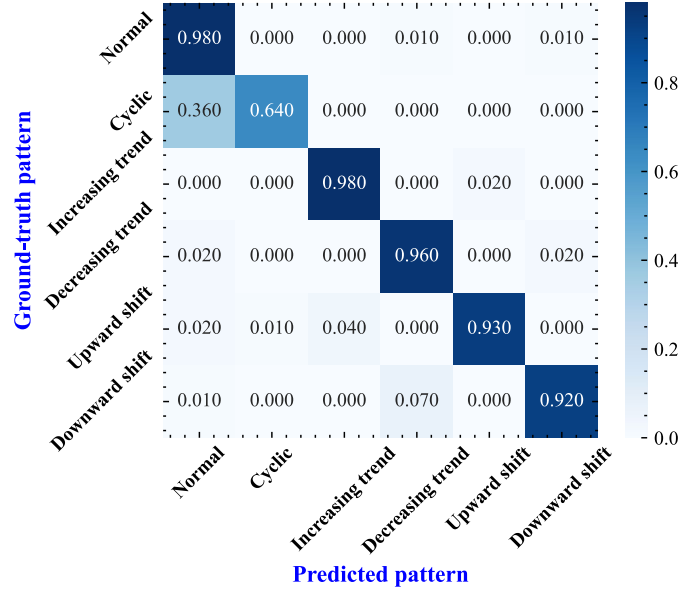


Fig. 12 Classification result on SCCTS data.

5.3 Scalability test and feature visualization

The scalability and robustness of the AI model is important for deployment in real industrial scenarios. Scalability indicates the capacity of the proposed method to process massive data. In this part, a variable-control experiment is conducted to investigate the changes in computational efficiency of DMP-MDNet when the data size varies. A baseline three-layer RBF network (RBFNN) was also implemented for comparison. The RBFNN adopted conventional hand-crafted feature-based learning paradigm where six time-domain features are taken as inputs, including the mean, standard deviation, maximum, minimum, energy, and entropy.

The two models were initially trained using the SCCTS dataset with the same training parameters. DMP-MDNet converged after just five epochs, and its training accuracy stabilized at 99.85%. In contrast, RBFNN achieved an accuracy of only 70% after 100 epochs. This stark contrast highlights the significant impact of feature quality on the performance of conventional methods relying on hand-crafted features. We then performed t-SNE with perplexity 200 to visualize the raw SCCTS samples as well as features extracted by RBFNN and DMP-MDNet. The t-SNE plots are presented in **Fig. 13**. The hand-crafted features used in RBFNN are heavily mixed across different categories, while the data representations obtained through DMP show a much more conspicuous clustering structure with little overlap. It is evident that the six time-domain features were insufficient to adequately represent the comprehensive information in the raw data.

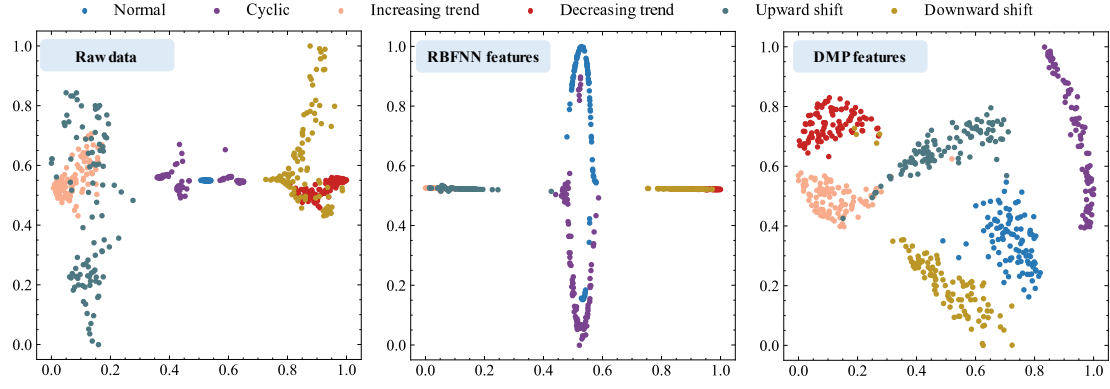
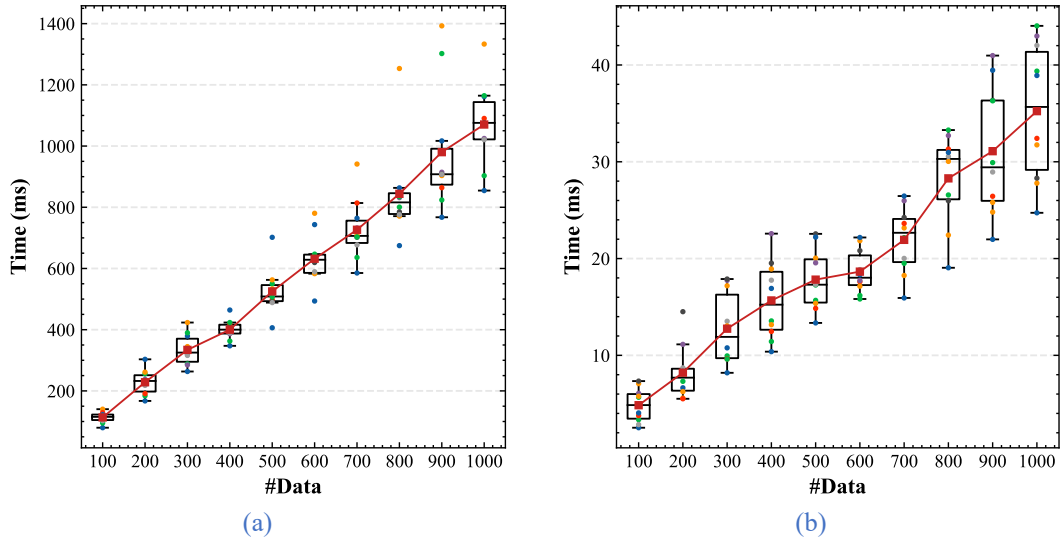


Fig. 13 t-SNE visualization of raw SCCTS data and extracted features.

The computational time of these two pre-trained models on the test sets with varying sizes was then compared. Ten runs were made for each model. As depicted in **Fig. 14** (c) and (d), DMP-MDNet exhibits a slower prediction time compared to the RBFNN, primarily due to its sophisticated network structure for improved feature extraction capacity. Nevertheless, it is noteworthy in **Fig. 14** (a) and (b) that the DMP encoding process is significantly faster than the feature extraction operation performed by the RBFNN. DMP-MDNet took about only 35.2 milliseconds to transform 1000 samples with lengths of 60 into DMP matrices. In contrary, RBFNN costed 1070.4 milliseconds to complete the feature extraction process for the same samples. Additionally, it was observed that the time required for traditional feature extraction increased substantially as the dataset size grew, while the DMP calculation time remained relatively stable. This leads to the conclusion that DMP-MDNet offers certain advantages in terms of computational efficiency.



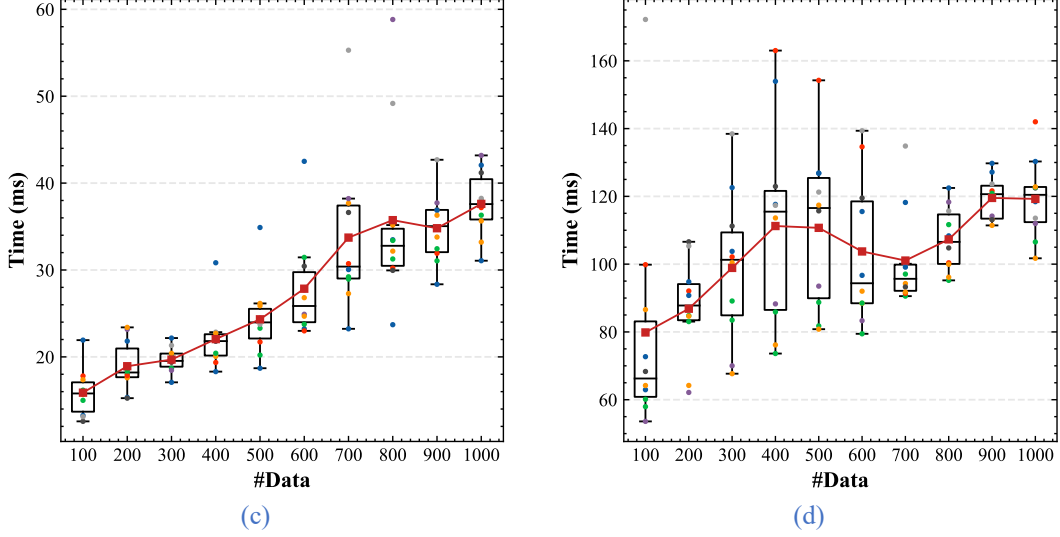


Fig. 14 Changes in computational time with data size. (a) Feature extraction time for RBFNN; (b) Feature representation for DMP-MDNet; (c) Prediction time for RBFNN; (d) Prediction time for DMP-MDNet.

The computational complexity of DMP-MDNet is further discussed from the theoretical perspective. The time complexity of the DMP module mainly arises from the iterations of instructions to calculate the similarities between subsequences. The time required to compute the similarity between two subsequences depends on the time to calculate their dot product. The dot product of $(j + 1)^{th}$ and $(i + 1)^{th}$ segments can be calculated in constant time based on the original dot product of j^{th} subsequence and i^{th} subsequence as shown in Eq. (11). Therefore, the time complexity of DMP transformation for a sequence with the length of n can be approximated to $O(n)$, which is more efficient than that of the hand-crafted feature extraction process in RBFNN with $O(n^2)$.

$$s_{j+1} \cdot s_{i+1} = \sum_{k=0}^{w-1} (x_{t_{i+1+k}} * x_{t_{j+1+k}}) = s_j \cdot s_i - x_{t_i} * x_{t_j} + x_{t_{j+w}} * x_{t_{i+w}} \quad (11)$$

The complexity of MDNet is measured by floating-point operations (FLOPs) and the total number of model parameters, two metrics commonly used to estimate the time and space complexity of deep learning models. Reduced FLOPs and parameters indicate lower model complexity, which theoretically mitigates the difficulty to handle large-scale input. As shown in **Table 1**, MDNet has 952,350 parameters and FLOPs of about 0.056G when the input size is $16 \times 30 \times 30$. In comparison, with the same input size, MDNet's variant without dilated convolution has more parameters and FLOPs, increasing by 3.42% and 103.57%, respectively. This suggests the enhanced computational efficiency of the proposed MDNet.

Table 1 Model complexity

Network	# Parameters	# FLOPs
MDNet without dilated convolution	984,926	0.114G
MDNet	952,350	0.056G

5.4 Model comparison

25 UCR datasets from various domains were employed to investigate the classification accuracy of DMP-MDNet. The results were compared to state-of-the-art (SOTA) methods. Note that the criteria for categorizing different classes within these datasets were not uniform. For instance, as illustrated in **Fig. 15**, the Car and Beef datasets fall into the category of TSC problems where local patterns play a predominant role. Conversely, datasets such as CBF and ECG5000 exhibit significant differences in overall data trends, highlighting a global pattern-dominated problem.

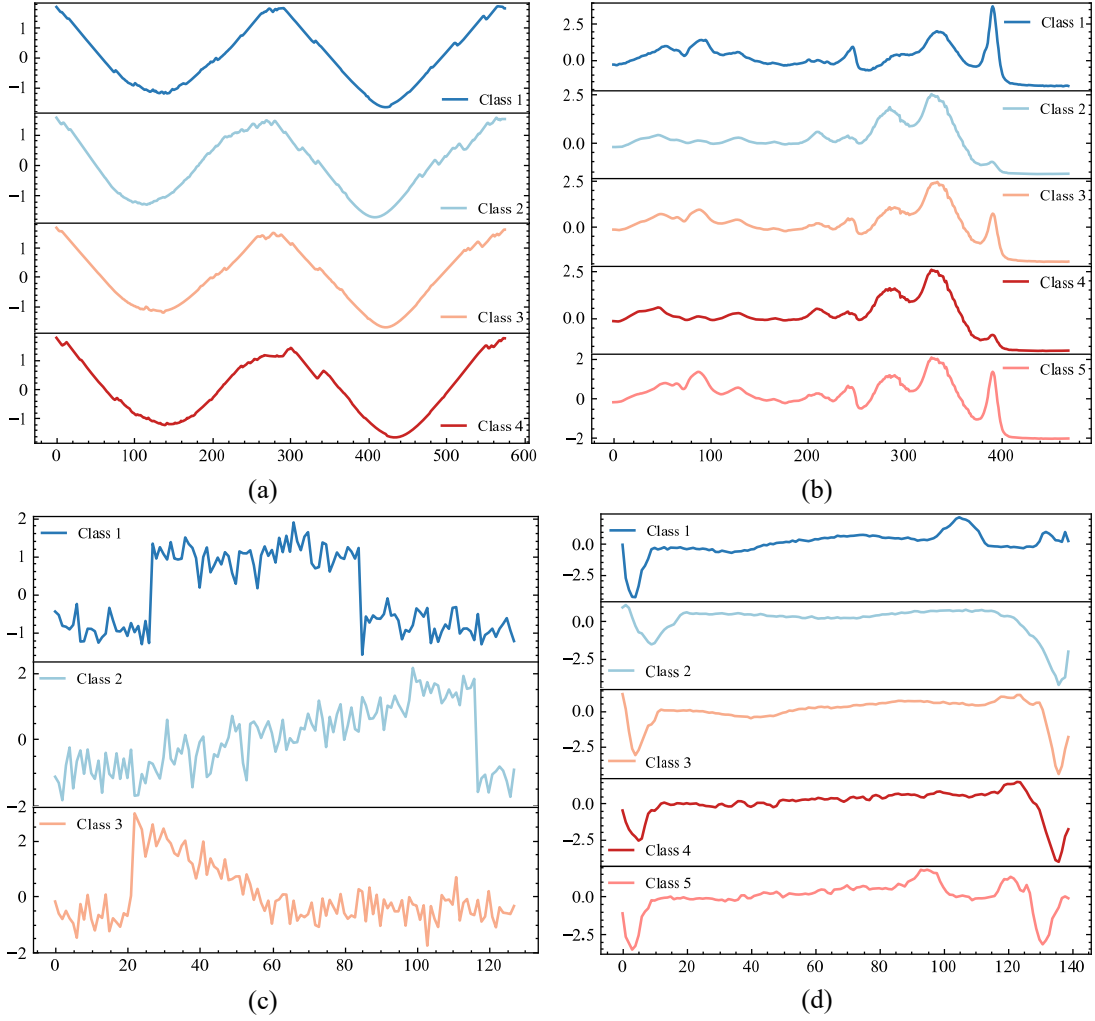


Fig. 15 Dataset visualization. (a) Car; (b) Beef; (c) CBF; (d) ECG5000;

14 SOTA methods, implemented using the Aeon toolkit [70], were selected for comparison. Based on the taxonomy of TSC methods described in Section 2.2, these methods can be categorized as follows:

1) Distance-based methods include ShapeDTW [39] and KNN.

2) Nine methods belonging to the hand-crafted feature-based category: BOSS [71], cBOSS [45], and WEASEL [44] perform classification according to symbolic features. STC [42] is based on shapelets, while TSF [43] and RISE [51] focus on interval features. TSFresh [46] employs feature engineering to extract statistical features. Hive-Cote [62] and Catch22 [52] perform ensemble learning based on multiple features.

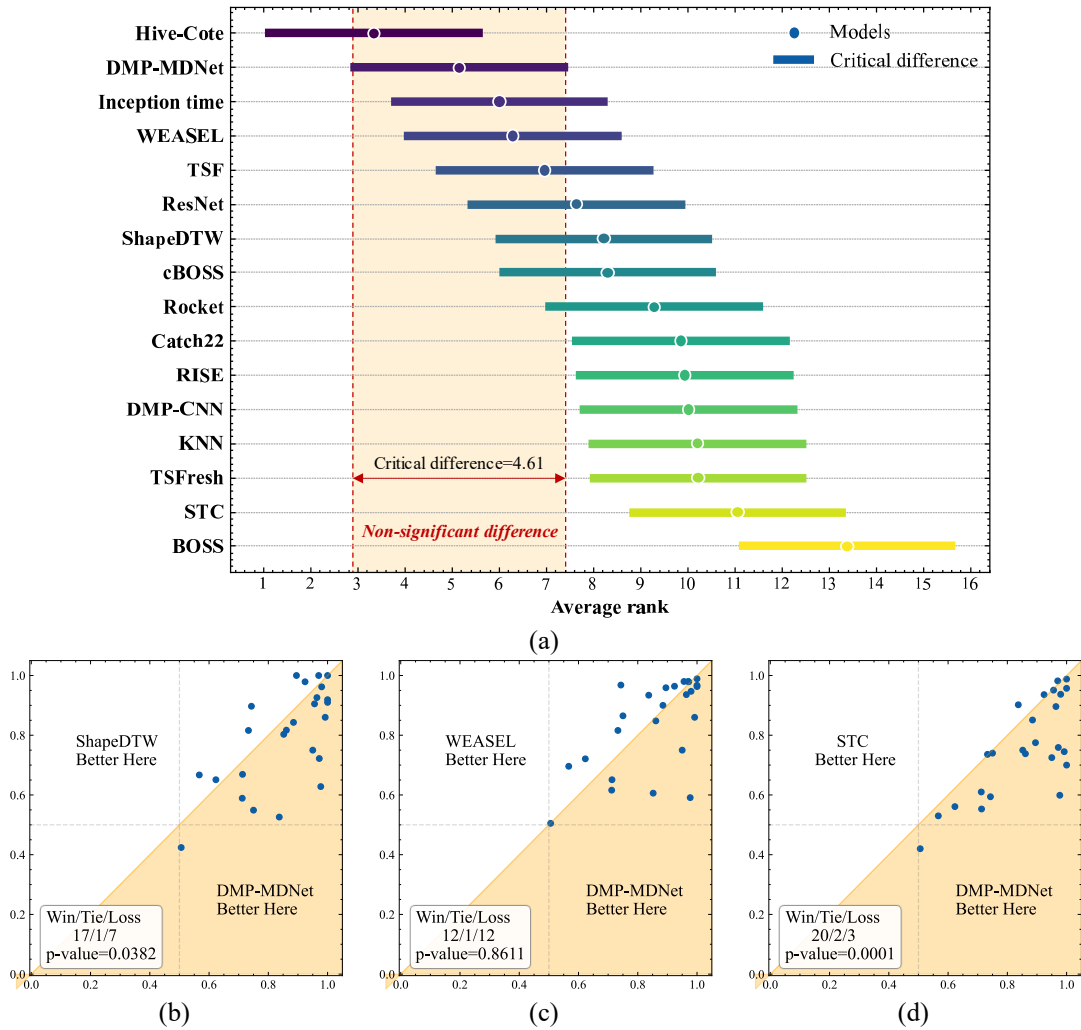
3) Three deep feature-based methods: ResNet [72] extracts features through a residual module with skip connections. Inception time [54] performs temporal feature extraction through multiple connected inception modules. Rocket [53] employs multiple random convolution kernels for feature transformation.

In addition, a baseline method, DMP-CNN, was constructed to investigate the effects of various receptive fields on DMP-MDNet. DMP-CNN retained only the middle receptive field branch in **Fig. 6**.

The training epochs for both DMP-MDNet and DMP-CNN were set to 500, and the learning rate was fixed at 0.001, with provisions for early stopping and learning rate decay. The batch size and DMP

size were dynamically adjusted based on the series length and dataset size. Typically, the DMP size was set to approximately 70% of the sequence length, considering the graphics memory constraints. The performance metric utilized the average test accuracy over 10 runs for each classifier.

DMP-MDNet achieved the highest accuracy in eight datasets, as shown in Appendix A, which is on par with Hive-Cote. This indicates that DMP-MDNet is applicable across various application areas. A Friedman–Nemenyi test was conducted, and the results are visualized in a critical difference diagram in Fig. 16(a). DMP-MDNet ranks second among all models, slightly behind Hive-Cote. This is reasonable as DMP-MDNet does not integrate multiple classifiers with diverse advantages as Hive-Cote does. Critical pairwise comparisons were also made between DMP-MDNet and the best models from different TSC types, and the win/tie/loss diagrams with a tolerance of ± 0.005 are provided in Fig. 16(b)–Fig. 16(h). DMP-MDNet significantly outperformed the best distance-based model ShapeDTW, the best shapelet feature-based model STC, the best interval-based model TSF, and the best feature engineering-based model TSFresh. The p-value obtained during the Friedman–Nemenyi test is indicated in the lower-left corner. A larger p-value suggests a less significant difference in model performance. Clearly, DMP-MDNet exhibits classification ability similar to WEASEL and Inception time. Moreover, as shown in Fig. 16(i), DMP-MDNet outperforms DMP-CNN, indicating the utility of the multi-receptive field structure.



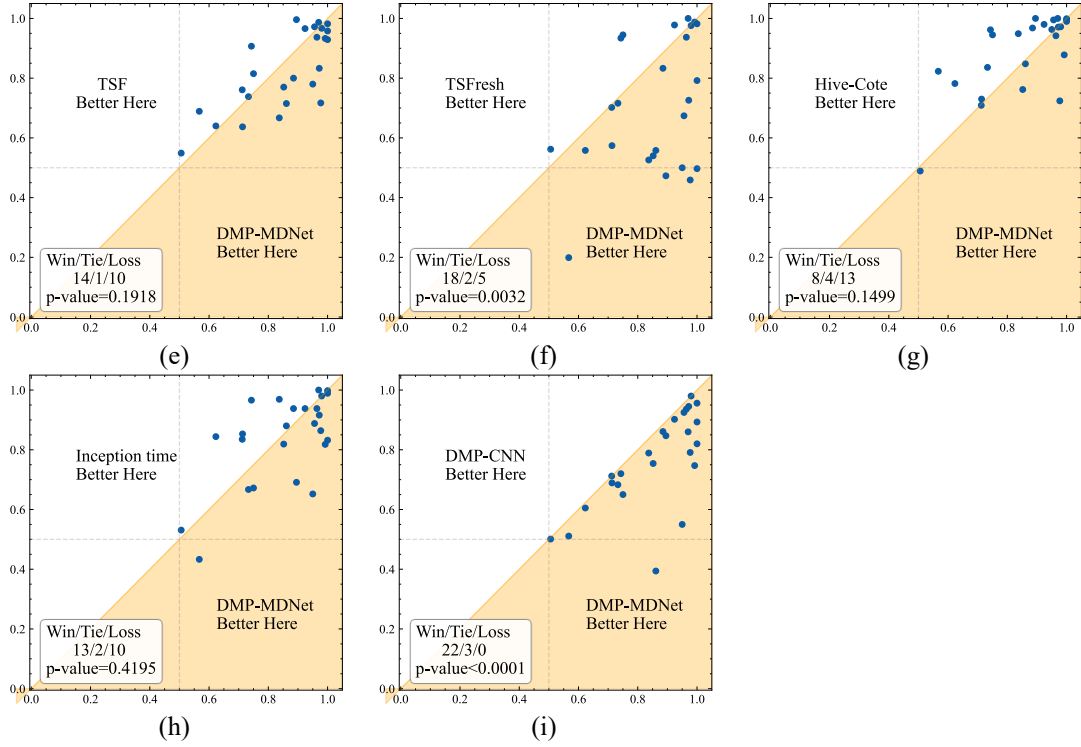


Fig. 16 Model comparison. (a) Friedman–Nemenyi test results; (b)–(i) Critical pairwise comparisons between classifiers.

It is important to highlight that DMP-MDNet excels in tackling TSC problems, where distinctions among different categories rely on overall trends or shapes. However, there is still a gap between DMP-MDNet and classifiers that emphasize local patterns. One potential reason for this difference is the fixed window size used in DMP calculation, which limits its ability to capture similarity information within a narrow range. For instance, in the case of the Car dataset, its local features can be overlooked by larger windows, potentially leading to similarity in DMP matrices among different categories.

5.5 Discussion

The experimental results indicate that DMP-MDNet excels in recognition performance on both real BIW data and public datasets. This method offers several advantages:

1) Improved temporal representations: DMP operates directly on raw data sequences without the need for expert knowledge, preserving data information more effectively. It signifies a shift of manufacturing process monitoring from traditional experience-based monitoring to data-driven, automated abnormal pattern recognition. Additionally, the similarity metric used in DMP is robust to sequence magnitude and length, enhancing the generalizability and robustness of the proposed temporal representation DMP across diverse tasks. The multi-receptive field structure of MDNet further realizes multi-granularity information fusion. Classification effect is thus improved by effectively preserving raw data information and leveraging representations with superior discriminative and generalization capabilities.

2) Improved speed: DMP transformation is computationally efficient, outperforming traditional feature extraction methods. This efficiency is attributed to the use of STOMP algorithm [73], which has a time complexity of $O(n)$. Moreover, dilated convolution in MDNet ensures efficient inference.

However, it is important to acknowledge some limitations: First, erroneous data may still be present

in NOP samples in real production data, as our method does not include denoising operations. Second, while DMP-MDNet performed well on TSC problems dominated by global patterns, its performance was insufficient for TSC problems dominated by local patterns. Therefore, we plan to explore the integration of attention mechanisms in the future: Time-domain attention is used to enhance the capture of local features. Frequency-domain attention offers an adaptive denoising capability.

6. Conclusion

This study presents an innovative multi-scale TSC-based method to identify multi-scale abnormal patterns stemming from various anomalous manufacturing conditions. Long-period production sequences are sliced and classified at multiple observation scales to investigate both global trends and local properties, so as to analyze the dynamics of production status throughout the overall manufacturing process. The classification is performed by a TSC model DMP-MDNet, which is designed to improve the robustness and effectiveness of manufacturing pattern recognition. This model leverages a temporal representation, DMP, to encode the generalizable scale- and magnitude-invariant properties of time-series data. Additionally, MDNet, a multi-branch network featuring dilated convolution and the fusion of multiple receptive field sizes is introduced to exploit multi-granularity information. A comprehensive case study on BIW production is introduced, in addition to conducting extensive experiments on 25 public datasets. The primary conclusions are drawn as follows.

1) Without the need for domain knowledge to guide tailored feature extraction, the proposed DMP successfully promotes the enhanced quality of time-series representation. The DMP matrix proves discriminative and generalizable by encoding scale- and magnitude-invariant features. When the model pre-trained by the BIW production dataset is applied directly to a control chart dataset from other sources, the classification accuracy shows a trivial decrease of about 2%.

2) The proposed classification module MDNet improves TSC effectiveness by integrating multiple receptive field sizes. Thanks to its ability to capture and model multi-granular valuable information within the input data, DMP-MDNet significantly outperforms its variant with only a single receptive field size on 25 public datasets, with an average increase of 15.29% in terms of accuracy.

3) With data representation with finer quality, DMP-MDNet surpasses various SOTA methods. It achieves the best accuracy 8 times, sharing first place with HIVE-COTE but with a much more concise model structure. In particular, DMP-MDNet excels in global pattern-dominated tasks, with the highest improvement of 13.07% in classification accuracy.

4) Compared with the conventional hand-crafted feature-based classifier, DMP-MDNet exhibits enhanced scalability attributed to the $O(n)$ complexity of the DMP encoding process and the utilization of dilated convolution.

Experimental results demonstrate that DMP-MDNet not only achieves high accuracy but also eliminates the need for labor-intensive feature engineering while exhibiting stronger generalization capabilities. The introduction of the novel TSC method provides valuable insights into multi-scale anomalies within complete manufacturing processes. It stands as a smart and robust analytical tool capable of harnessing the vast amounts of production data generated in the realm of cyber manufacturing. Future research endeavors should consider attention-enhanced models and delve into multivariate process monitoring for further advancements in this field.

629

Acknowledgements

630

This work is sponsored by the National Natural Science Foundation of China [No.52075338], and
631 Shanghai Key lab of Advanced Manufacturing Environment.

632

Disclosure statement

633

No potential conflict of interest was reported by the author(s).

634

Data availability statements

635

The BIW production data are available from the corresponding author upon request. Public datasets in
636 Section 5 are available in https://www.cs.ucr.edu/~eamonn/time_series_data_2018/.

637

Appendix A. Accuracies of different classifiers over 25 UCR datasets

Dataset	State-of-the-art methods												This paper			
	ShapeDT W	KNN	Catch22	Hive- Cote	RISE	TSF	TSFRESH	STC	BOSS	cBOSS	WEASE L	Rocket	Inception time	ResNet	DMP- CNN	DMP- MDNet
ArrowHead	0.817	0.799	0.680	0.848	0.727	0.715	0.558	0.738	0.542	0.783	0.848	0.527	0.880	0.314	0.394	0.861
Beef	0.667	0.667	0.536	0.823	0.770	0.689	0.199	0.530	0.399	0.606	0.696	0.416	0.433	0.538	0.511	0.567
BeetleFly	0.750	0.750	0.735	0.963	0.745	0.780	0.500	0.725	0.700	0.815	0.750	0.749	0.652	0.853	0.550	0.950
BirdChicken	0.549	0.549	0.845	0.945	0.775	0.815	0.945	0.740	0.849	0.920	0.865	0.715	0.672	0.851	0.650	0.750
BME	0.860	0.826	0.848	0.878	0.721	0.933	0.988	0.745	0.553	0.683	0.860	0.974	0.818	0.943	0.747	0.992
Car	0.816	0.733	0.718	0.836	0.733	0.738	0.716	0.736	0.583	0.775	0.816	0.406	0.667	0.467	0.683	0.733
CBF	0.905	0.852	0.938	0.995	0.836	0.972	0.674	0.951	0.828	0.959	0.980	0.994	0.888	0.748	0.925	0.956
Chinatown	0.962	0.944	0.871	0.972	0.941	0.967	0.976	0.937	0.813	0.888	0.947	0.967	0.980	0.831	0.980	0.980
ChlorineCon	0.628	0.650	0.572	0.724	0.739	0.717	0.459	0.599	0.539	0.630	0.591	0.496	0.864	0.819	0.791	0.977
CinCECGTorso	0.910	0.897	0.785	0.990	0.928	0.958	0.792	0.700	0.476	0.679	0.967	0.407	0.832	0.768	0.820	1.000
Coffe	1.000	1.000	0.992	0.996	0.982	0.982	0.982	0.957	0.857	0.978	0.989	0.807	0.998	0.536	0.893	1.000
CricketX	0.669	0.576	0.553	0.730	0.519	0.637	0.574	0.553	0.632	0.597	0.651	0.697	0.853	0.808	0.689	0.713
CricketY	0.651	0.566	0.515	0.782	0.523	0.640	0.558	0.561	0.654	0.662	0.721	0.687	0.844	0.792	0.605	0.623
ECG5000	0.926	0.924	0.932	0.942	0.933	0.937	0.937	0.896	0.857	0.937	0.936	0.943	0.938	0.937	0.936	0.964
ECGFiveDays	0.919	0.796	0.748	0.999	0.989	0.929	0.497	0.988	0.707	0.993	0.963	0.962	0.989	0.727	0.956	1.000
IEPGST	1.000	1.000	0.981	1.000	0.933	0.996	0.473	0.775	0.967	0.911	0.959	0.890	0.691	0.878	0.847	0.895
Lightning2	0.803	0.754	0.734	0.762	0.711	0.770	0.540	0.750	0.540	0.718	0.606	0.627	0.819	0.721	0.754	0.852
Lightning7	0.589	0.575	0.642	0.709	0.598	0.761	0.702	0.610	0.452	0.604	0.616	0.713	0.835	0.810	0.712	0.712
MSRT	0.897	0.897	0.906	0.962	0.895	0.907	0.934	0.594	0.694	0.872	0.968	0.548	0.966	0.952	0.720	0.743
RefriDevices	0.424	0.394	0.513	0.489	0.488	0.549	0.562	0.420	0.434	0.495	0.505	0.503	0.531	0.512	0.501	0.506
StarLightCurves	0.979	0.848	0.967	0.980	0.972	0.966	0.978	0.936	0.922	0.956	0.964	0.837	0.938	0.968	0.902	0.924
ToeSegment1	0.526	0.679	0.811	0.949	0.766	0.667	0.526	0.902	0.632	0.867	0.934	0.904	0.969	0.954	0.789	0.837
ToeSegment2	0.843	0.807	0.781	0.968	0.855	0.800	0.833	0.851	0.831	0.950	0.900	0.866	0.938	0.884	0.861	0.885
Trace	1.000	0.759	0.996	1.000	0.936	0.987	1.000	0.982	0.710	0.998	0.980	1.000	1.000	1.000	0.945	0.970
UMD	0.722	0.805	0.866	0.971	0.588	0.833	0.726	0.759	0.562	0.788	0.979	0.932	0.916	0.953	0.861	0.972
Winning times	3	2	0	9	0	0	3	0	0	0	2	1	7	1	1	8

References

- [1] Dai W, Jiao Y, Erdt M, Sourin A. Self-supervised Pairing Image Clustering and Its Application in Cyber Manufacturing. *Proc. - 2020 Int. Conf. Cyberworlds, CW 2020*, 2020, p. 25–32. <https://doi.org/10.1109/CW49994.2020.00012>.
- [2] Maged A, Lui CF, Haridy S, Xie M. Variational AutoEncoders-LSTM based fault detection of time-dependent high dimensional processes. *Int J Prod Res* 2023;62:1092–107. <https://doi.org/10.1080/00207543.2023.2175591>.
- [3] Lee J, Bagheri B, Jin C. Introduction to cyber manufacturing. *Manuf Lett* 2016;8:11–5. <https://doi.org/10.1016/j.mfglet.2016.05.002>.
- [4] Su Y, Liu R, Zhao Y, Sun W, Niu C, Pei D. Robust anomaly detection for multivariate time series through stochastic recurrent neural network. *Proc ACM SIGKDD Int Conf Knowl Discov Data Min* 2019;1485:2828–37. <https://doi.org/10.1145/3292500.3330672>.
- [5] Kao LJ, Chiu CC. Application of integrated recurrent neural network with multivariate adaptive regression splines on SPC-EPC process. *J Manuf Syst* 2020;57:109–18. <https://doi.org/10.1016/j.jmsy.2020.07.020>.
- [6] Zan T, Liu Z, Wang H, Wang M, Gao X. Control chart pattern recognition using the convolutional neural network. *J Intell Manuf* 2020;31:703–16. <https://doi.org/10.1007/s10845-019-01473-0>.
- [7] Giannetti C, Essien A. Towards scalable and reusable predictive models for cyber twins in manufacturing systems. *J Intell Manuf* 2022;33:441–55. <https://doi.org/10.1007/s10845-021-01804-0>.
- [8] Liu C, Chen K, Jin S, Qu Y. Variation Pattern Recognition of the BIW OCMM Online Measurement Data Based on LSTM NN. *IEEE Access* 2019;7:69007–14. <https://doi.org/10.1109/ACCESS.2019.2918784>.
- [9] Huang JW, Lee PJ, Jaysawal BP. Multiscale Control Chart Pattern Recognition Using Histogram-Based Representation of Value and Zero-Crossing Rate. *IEEE Trans Ind Electron* 2022;69:684–93. <https://doi.org/10.1109/TIE.2021.3050355>.
- [10] Farahani MA, McCormick MR, Gianinny R, Hudachek F, Harik R, Liu Z, et al. Time-series pattern recognition in Smart Manufacturing Systems: A literature review and ontology. *J Manuf Syst* 2023;69:208–41. <https://doi.org/10.1016/j.jmsy.2023.05.025>.
- [11] Guo S, Guo W. Process Monitoring and Fault Prediction in Multivariate Time Series Using Bag-of-Words. *IEEE Trans Autom Sci Eng* 2022;19:230–42. <https://doi.org/10.1109/TASE.2020.3026065>.
- [12] Shewhart WA. Economic Quality Control of Manufactured Product. *Bell Syst Tech J* 1930;9:364–89. <https://doi.org/10.1002/j.1538-7305.1930.tb00373.x>.
- [13] Page ES. Continuous Inspection Schemes. *Biometrika* 1954;41:100. <https://doi.org/10.2307/2333009>.
- [14] Roberts SW. Control Chart Tests Based on Geometric Moving Averages. *Technometrics* 2000;42:97. <https://doi.org/10.2307/1271439>.
- [15] Li W, Zhang C, Tsung F, Mei Y. Nonparametric monitoring of multivariate data via KNN learning. *Int J Prod Res* 2021;59:6311–26. <https://doi.org/10.1080/00207543.2020.1812750>.
- [16] He S, Jiang W, Deng H. A distance-based control chart for monitoring multivariate processes using support vector machines. *Ann Oper Res* 2018;263:191–207.

- <https://doi.org/10.1007/s10479-016-2186-4>.
- [17] Yu J, Li S, Liu X, Gao Y, Wang S, Liu C. Dynamic convolutional gated recurrent unit attention auto-encoder for feature learning and fault detection in dynamic industrial processes. *Int J Prod Res* 2023;61:7434–52. <https://doi.org/10.1080/00207543.2022.2149874>.
 - [18] Hachicha W, Ghorbel A. A survey of control-chart pattern-recognition literature (1991-2010) based on a new conceptual classification scheme. *Comput Ind Eng* 2012;63:204–22. <https://doi.org/10.1016/j.cie.2012.03.002>.
 - [19] Psarommatis F, May G, Azamfirei V. Envisioning maintenance 5.0: Insights from a systematic literature review of Industry 4.0 and a proposed framework. *J Manuf Syst* 2023;68:376–99. <https://doi.org/10.1016/j.jmsy.2023.04.009>.
 - [20] Ghaleb M, Taghipour S. Evidence-based study of the impacts of maintenance practices on asset sustainability. *Int J Prod Res* 2023;61:8719–50. <https://doi.org/10.1080/00207543.2022.2152893>.
 - [21] Singh J, Ahuja IPS, Singh H, Singh A. Development and Implementation of Autonomous Quality Management System (AQMS) in an Automotive Manufacturing using Quality 4.0 Concept– A Case Study. *Comput Ind Eng* 2022;168:108121. <https://doi.org/10.1016/j.cie.2022.108121>.
 - [22] Teiwes H, Rössinger M, Gonter M, Herrmann C, Thiede S. Data based analysis and improvement of energy efficiency in the automotive body shop. *J Clean Prod* 2021;284. <https://doi.org/10.1016/j.jclepro.2020.125269>.
 - [23] Sari E, Ma'aram A, Shaharoun AM, Chofreh AG, Goni FA, Klemeš JJ, et al. Measuring sustainable cleaner maintenance hierarchical contributions of the car manufacturing industry. *J Clean Prod* 2021;312. <https://doi.org/10.1016/j.jclepro.2021.127717>.
 - [24] Bucay-Valdiviezo J, Escudero-Villa P, Paredes-Fierro J, Ayala-Chauvin M. Leveraging Classical Statistical Methods for Sustainable Maintenance in Automotive Assembly Equipment. *Sustainability* 2023;15:15604. <https://doi.org/10.3390/su152115604>.
 - [25] Antosz K, Jasiulewicz-Kaczmarek M, Waszkowski R, Machado J. Application of Lean Six Sigma for sustainable maintenance: case study. *IFAC-PapersOnLine*, vol. 55, 2022, p. 181–6. <https://doi.org/10.1016/j.ifacol.2022.09.204>.
 - [26] Fingerhut F, Harsha C, Eghbalian A, Jacobs T, Tabassian M, Verbeke R, et al. Data-Driven Usage Profiling and Anomaly Detection in Support of Sustainable Machining Processes. *IEEE Int. Conf. Data Min. Work. ICDMW*, IEEE; 2022, p. 127–36. <https://doi.org/10.1109/ICDMW58026.2022.00026>.
 - [27] Gómez ÁLP, Maimó LF, Celdrán AH, Clemente FJG. SUSAN: A Deep Learning based anomaly detection framework for sustainable industry. *Sustain Comput Informatics Syst* 2023;37:100842. <https://doi.org/10.1016/j.suscom.2022.100842>.
 - [28] Xia T, Dong Y, Xiao L, Du S, Pan E, Xi L. Recent advances in prognostics and health management for advanced manufacturing paradigms. *Reliab Eng Syst Saf* 2018;178:255–68. <https://doi.org/10.1016/j.ress.2018.06.021>.
 - [29] Perera YS, Ratnaweera DAAC, Dasanayaka CH, Abeykoon C. The role of artificial intelligence-driven soft sensors in advanced sustainable process industries: A critical review. *Eng Appl Artif Intell* 2023;121:105988. <https://doi.org/10.1016/j.engappai.2023.105988>.
 - [30] Lee H. Framework and development of fault detection classification using IoT device and cloud environment. *J Manuf Syst* 2017;43:257–70. <https://doi.org/10.1016/j.jmsy.2017.02.007>.
 - [31] Liu C, Chen K, Jin S, Qu Y, Yu J, Zhou B. An integrated method for variation pattern recognition of BIW OCMM online measurement data. *Int J Prod Res* 2022;60:1932–53.

- <https://doi.org/10.1080/00207543.2021.1877841>.
- [32] Liao Y, Ragai I, Huang Z, Kerner S. Manufacturing process monitoring using time-frequency representation and transfer learning of deep neural networks. *J Manuf Process* 2021;68:231–48. <https://doi.org/10.1016/j.jmapro.2021.05.046>.
- [33] Mahato V, Obeidi MA, Brabazon D, Cunningham P. Detecting voids in 3D printing using melt pool time series data. *J Intell Manuf* 2022;33:845–52. <https://doi.org/10.1007/s10845-020-01694-8>.
- [34] Qian J, Wei C, Song Z. A Novel Sequence Discriminative Feature Extraction Network and Its Application in Offline Industrial Fault Pattern Clustering. *IEEE Trans Ind Informatics* 2024;20:3025–33. <https://doi.org/10.1109/TII.2023.3301045>.
- [35] Yousef I, Tulsyan A, Shah SL, Gopaluni RB. Visual analytics for process monitoring: Leveraging time-series imaging for enhanced interpretability. *J Process Control* 2023;132:103127. <https://doi.org/10.1016/j.jprocont.2023.103127>.
- [36] Zhao S, Duan Y, Roy N, Zhang B. A deep learning methodology based on adaptive multiscale CNN and enhanced highway LSTM for industrial process fault diagnosis. *Reliab Eng Syst Saf* 2024;249:110208. <https://doi.org/10.1016/j.ress.2024.110208>.
- [37] Mori U, Mendiburu A, Keogh E, Lozano JA. Reliable early classification of time series based on discriminating the classes over time. *Data Min Knowl Discov* 2017;31:233–63. <https://doi.org/10.1007/s10618-016-0462-1>.
- [38] Li H, Liu J, Yang Z, Liu RW, Wu K, Wan Y. Adaptively constrained dynamic time warping for time series classification and clustering. *Inf Sci (Ny)* 2020;534:97–116. <https://doi.org/10.1016/j.ins.2020.04.009>.
- [39] Zhao J, Itti L. shapeDTW: Shape Dynamic Time Warping. *Pattern Recognit* 2018;74:171–84. <https://doi.org/10.1016/j.patcog.2017.09.020>.
- [40] Chen L, Ng R. On The Marriage of Lp-norms and Edit Distance. *Proc. 2004 VLDB Conf. 30th Int. Conf. Very Large Databases, VLDB Endowment; 2004, p. 792–803*. <https://doi.org/10.1016/B978-012088469-8.50070-X>.
- [41] Chen L, Özsu MT, Oria V. Robust and fast similarity search for moving object trajectories. *Proc. ACM SIGMOD Int. Conf. Manag. Data, New York, NY, USA: Association for Computing Machinery; 2005, p. 491–502*. <https://doi.org/10.1145/1066157.1066213>.
- [42] Hills J, Lines J, Baranauskas E, Mapp J, Bagnall A. Classification of time series by shapelet transformation. *Data Min Knowl Discov* 2014;28:851–81. <https://doi.org/10.1007/s10618-013-0322-1>.
- [43] Li G, Choi B, Xu J, Bhowmick SS, Chun KP, Wong GLH. Efficient Shapelet Discovery for Time Series Classification. *IEEE Trans Knowl Data Eng* 2022;34:1149–63. <https://doi.org/10.1109/TKDE.2020.2995870>.
- [44] Schäfer P, Leser U. Fast and accurate time series classification with WEASEL. *Int Conf Inf Knowl Manag Proc* 2017;Part F1318:637–46. <https://doi.org/10.1145/3132847.3132980>.
- [45] Middlehurst M, Vickers W, Bagnall A. Scalable dictionary classifiers for time series classification. vol. 11871 LNCIS. Springer International Publishing; 2019. https://doi.org/10.1007/978-3-030-33607-3_2.
- [46] Christ M, Braun N, Neuffer J, Kempa-Liehr AW. Time Series Feature Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package). *Neurocomputing* 2018;307:72–7. <https://doi.org/10.1016/j.neucom.2018.03.067>.

- [47] Aydin İ, Karakose M, Akin E. A new method for time series classification using multi-dimensional phase space and a statistical control chart. *Neural Comput Appl* 2020;32:7439–53. <https://doi.org/10.1007/s00521-019-04270-1>.
- [48] Medina R, Macancela JC, Lucero P, Cabrera D, Sánchez RV, Cerrada M. Gear and bearing fault classification under different load and speed by using Poincaré plot features and SVM. *J Intell Manuf* 2022;33:1031–55. <https://doi.org/10.1007/s10845-020-01712-9>.
- [49] Zhang Y, Hou Y, OuYang K, Zhou S. Multi-scale signed recurrence plot based time series classification using inception architectural networks. *Pattern Recognit* 2022;123. <https://doi.org/10.1016/j.patcog.2021.108385>.
- [50] Deng H, Runger G, Tuv E, Vladimir M. A time series forest for classification and feature extraction. *Inf Sci (Ny)* 2013;239:142–53. <https://doi.org/10.1016/j.ins.2013.02.030>.
- [51] Bagnall A, Lines J, Hills J, Bostrom A. Time-Series Classification with COTE: The Collective of Transformation-Based Ensembles. *IEEE Trans Knowl Data Eng* 2015;27:2522–35. <https://doi.org/10.1109/TKDE.2015.2416723>.
- [52] Lubba CH, Sethi SS, Knaute P, Schultz SR, Fulcher BD, Jones NS. catch22: CAnonical Time-series CHaracteristics: Selected through highly comparative time-series analysis. *Data Min Knowl Discov* 2019;33:1821–52. <https://doi.org/10.1007/s10618-019-00647-x>.
- [53] Dempster A, Petitjean F, Webb GI. ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels. *Data Min Knowl Discov* 2020;34:1454–95. <https://doi.org/10.1007/s10618-020-00701-z>.
- [54] Ismail Fawaz H, Lucas B, Forestier G, Pelletier C, Schmidt DF, Weber J, et al. InceptionTime: Finding AlexNet for time series classification. *Data Min Knowl Discov* 2020;34:1936–62. <https://doi.org/10.1007/s10618-020-00710-y>.
- [55] Karim F, Majumdar S, Darabi H, Chen S. LSTM Fully Convolutional Networks for Time Series Classification. *IEEE Access* 2017;6:1662–9. <https://doi.org/10.1109/ACCESS.2017.2779939>.
- [56] Ma Q, Li S, Cottrell GW. Adversarial Joint-Learning Recurrent Neural Network for Incomplete Time Series Classification. *IEEE Trans Pattern Anal Mach Intell* 2022;44:1765–76. <https://doi.org/10.1109/TPAMI.2020.3027975>.
- [57] Wu H, Triebe MJ, Sutherland JW. A transformer-based approach for novel fault detection and fault classification/diagnosis in manufacturing: A rotary system application. *J Manuf Syst* 2023;67:439–52. <https://doi.org/10.1016/j.jmsy.2023.02.018>.
- [58] Chen W, Shi K. Multi-scale Attention Convolutional Neural Network for time series classification. *Neural Networks* 2021;136:126–40. <https://doi.org/10.1016/j.neunet.2021.01.001>.
- [59] Lea C, Flynn MD, Vidal R, Reiter A, Hager GD. Temporal convolutional networks for action segmentation and detection. *Proc - 30th IEEE Conf Comput Vis Pattern Recognition, CVPR 2017 2017*;2017-Janua:1003–12. <https://doi.org/10.1109/CVPR.2017.113>.
- [60] Fullah Kamara A, Chen E, Liu Q, Pan Z. Combining contextual neural networks for time series classification. *Neurocomputing* 2020;384:57–66. <https://doi.org/10.1016/j.neucom.2019.10.113>.
- [61] Kenji Iwana B, Uchida S. Time series classification using local distance-based features in multi-modal fusion networks. *Pattern Recognit* 2020;97. <https://doi.org/10.1016/j.patcog.2019.107024>.
- [62] Lines J, Taylor S, Bagnall A. Time series classification with HIVE-COTE: The hierarchical vote collective of transformation-based ensembles. *ACM Trans Knowl Discov Data* 2018;12. <https://doi.org/10.1145/3182382>.
- [63] Ma G, Yang X, Zhang B, Shi Z. Multi-feature fusion deep networks. *Neurocomputing*

- 2016;218:164–71. <https://doi.org/10.1016/j.neucom.2016.08.059>.
- [64] Pham DT, Wani MA. Feature-based control chart pattern recognition. *Int J Prod Res* 1997;35:1875–90. <https://doi.org/10.1080/002075497194967>.
- [65] Kong J, Kim J, Bae J. HiFi-GAN: Generative adversarial networks for efficient and high fidelity speech synthesis. *Adv. Neural Inf. Process. Syst.*, vol. 2020- Decem, Red Hook, NY, USA: Curran Associates Inc.; 2020.
- [66] He Z, Cao Y, Du L, Xu B, Yang J, Cao Y, et al. MRFN: Multi-Receptive-Field Network for Fast and Accurate Single Image Super-Resolution. *IEEE Trans Multimed* 2020;22:1042–54. <https://doi.org/10.1109/TMM.2019.2937688>.
- [67] Zheng Y, Chen L, Lu X, Sen Y, Cheng H. Digital twin for geometric feature online inspection system of car body-in-white. *Int J Comput Integr Manuf* 2021;34:752–63. <https://doi.org/10.1080/0951192X.2020.1736637>.
- [68] Dau HA, Bagnall A, Kamgar K, Yeh CCM, Zhu Y, Gharghabi S, et al. The UCR time series archive. *IEEE/CAA J Autom Sin* 2019;6:1293–305. <https://doi.org/10.1109/JAS.2019.1911747>.
- [69] Gauri SK, Chakraborty S. Feature-based recognition of control chart patterns. *Comput Ind Eng* 2006;51:726–42. <https://doi.org/10.1016/j.cie.2006.07.013>.
- [70] Bagnall A, Lines J, Bostrom A, Large J, Keogh E. The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Min Knowl Discov* 2017;31:606–60. <https://doi.org/10.1007/s10618-016-0483-9>.
- [71] Schäfer P. The BOSS is concerned with time series classification in the presence of noise. *Data Min Knowl Discov* 2015;29:1505–30. <https://doi.org/10.1007/s10618-014-0377-7>.
- [72] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit* 2016;2016-Decem:770–8. <https://doi.org/10.1109/CVPR.2016.90>.
- [73] Zhu Y, Zimmerman Z, Senobari NS, Yeh CCM, Funning G, Mueen A, et al. Matrix profile II: Exploiting a novel algorithm and GPUs to break the one hundred million barrier for time series motifs and joins. *Proc - IEEE Int Conf Data Mining, ICDM* 2016;0:739–48. <https://doi.org/10.1109/ICDM.2016.126>.