Diffusion Language-Shapelets for Semi-supervised Time-Series Classification

Zhen Liu[1], **Wenbin Pei**[2,3], **Disen Lan**[1], **Qianli Ma**[1*]
[1]School of Computer Science and Engineering, South China University of Technology, Guangzhou, China
[2]School of Computer Science and Technology, Dalian University of Technology, Dalian, China
[3]Key Laboratory of Social Computing and Cognitive Intelligence (Dalian University of Technology), Ministry of Education
cszhenliu@mail.scut.edu.cn, peiwenbin@dlut.edu.cn, 202130480657@mail.scut.edu.cn, qianlima@scut.edu.cn

Abstract

Semi-supervised time-series classification could effectively alleviate the issue of lacking labeled data. However, existing approaches usually ignore model interpretability, making it difficult for humans to understand the principles behind the predictions of a model. Shapelets are a set of discriminative subsequences that show high interpretability in time series classification tasks. Shapelet learning-based methods have demonstrated promising classification performance. Unfortunately, without enough labeled data, the shapelets learned by existing methods are often poorly discriminative, and even dissimilar to any subsequence of the original time series. To address this issue, we propose the **Diff**usion Language-**Shape**lets model (**DiffShape**) for semi-supervised time series classification. In DiffShape, a self-supervised diffusion learning mechanism is designed, which uses real subsequences as a condition. This helps to increase the similarity between the learned shapelets and real subsequences by using a large amount of unlabeled data. Furthermore, we introduce a contrastive language-shapelets learning strategy that improves the discriminability of the learned shapelets by incorporating the natural language descriptions of the time series. Experiments have been conducted on the UCR time series archive, and the results reveal that the proposed DiffShape method achieves state-of-the-art performance and exhibits superior interpretability over baselines.

Introduction

Time series is a set of data points listed in chronological order, which is usually used to describe time-dependent phenomena, e.g., electrocardiograms (Maweu et al. 2021), electricity consumption (Cheng et al. 2020), and human activities (Chen et al. 2021). Recently, deep learning has been successfully applied to time series classification (TSC), mainly due to its powerful feature learning capability. In general, training deep models requires a large amount of labeled data. However, it is usually time-consuming and laborious to label time series data in many real-world applications. Semi-supervised classification (SSC) (Yang et al. 2022) allows using labeled and unlabeled data simultaneously for training, which could alleviate the issue of lacking labeled data.

SSC-based time series methods mainly apply consistency regularization and pseudo-labeling techniques. The former (Fan et al. 2021; Wei et al. 2023) uses sampled subsequences to design time series prediction loss and relation prediction loss, for learning temporal dependencies of labeled and unlabeled instances. The latter (Lee et al. 2013; Liu et al. 2023b) utilizes a model that predicts labels (pseudo-labels in reality) on unlabeled data for training. Although the aforementioned methods improve the classification performance of an SSC model, they ignore the interpretability of the model. As a result, it is difficult for humans to understand the predictions.

Shapelets are a set of discriminant subsequences (also called shapes) of time series (Ye and Keogh 2009), each of which is expected to represent a class optimally. Therefore, the use of shapelets can assist practitioners to well understand the meaning of time series, and expand the applications of TSC methods to some applications that are expected to have good model interpretability, e.g., medical diagnosis (Lin et al. 2019) and industrial safety (Yuan et al. 2020). However, existing shapelets learning-based methods (Grabocka et al. 2014; Li et al. 2021) usually rely on a large amount of labeled data. Unfortunately, in these applications, it is hard or expensive to label enough data for training. Without enough labeled data, the shapelets learned by the existing methods are often poorly discriminative, and even dissimilar to any subsequences of the time series.

Recently, diffusion models have achieved remarkable performance in time series prediction (Rasul et al. 2021) and imputation (Tashiro et al. 2021) tasks due to their effectiveness in generating samples. For instance, Shen and Kwok (2023) use the distribution of past observations to drive the diffusion model to generate moment values in the future. Shapelets learning aims to obtain partially discriminative subsequences based on the distribution of all subsequences of a time series, which inspires us to consider using the distribution of subsequences as a condition and then employing diffusion models to generate shapelets. However, in SSC, only a small number of subsequences contain class information. Thus, it is still a tough challenge to investigate how diffusion models can be used to generate shapelets that are conducive to enhancing classification performance.

In real life, humans can often rapidly identify instances belonging to the same class using just a small number of

[*]Qianli Ma is the corresponding author.

(even a few) labeled instances, along with their corresponding natural language descriptions. It is noteworthy that time series data has complex dynamic changes over time, often making it difficult to identify their class only through human intuition. Concurrently, Radford et al. (2021) employ image-text pairs for pre-training, showcasing the enhancement of image task performance through textual information. Recent studies have affirmed the effectiveness of leveraging text labels (Zhang et al. 2023) and large language models (Gruver et al. 2023; Jin et al. 2023) in the context of time series modeling. This naturally encourages us to take advantage of natural language descriptions to assist shapelet learning, thereby boosting classification accuracy and understanding.

In this paper, we propose the **Diff**usion Language-**Shape**lets model (**DiffShape**) for semi-supervised time series classification. Unlike most existing time series SSC methods, DiffShape automatically generates shapelets for each time series, improving the interpretability. Specifically, DiffShape incorporates two mechanisms. The first is the self-supervised learning mechanism based on the real subsequences as the diffusion condition, increasing the similarity between the generated shapelets and the original subsequences. The second is the contrastive language-shapelets learning mechanism, aimed at improving the discriminability of generated shapelets. By combining these mechanisms, DiffShape effectively leverages the text description of the time series and classification information of a classifier during training, making the generated shapelets more effective in enhancing classification performance.

The major contributions are summarized as follows:

- We propose the shapelet-based diffusion learning mechanism for semi-supervised time series classification. In particular, we employ the real subsequences of a large number of unlabeled instances as conditions in the diffusion process for self-supervised learning.
- We introduce the contrastive language-shapelets learning mechanism to alleviate the issue of lacking labeled data. By utilizing natural language descriptions generated from labels and pseudo-labels of time series, the discriminability of the generated shapelets is improved by aligning the distance between the transformed shapelet embeddings and their corresponding text embeddings.
- Extensive experiments on the UCR time series archive have been conducted, and the results show that the proposed DiffShape method outperforms existing time series SSC methods in terms of both classification performance and interoperability.

Related Work

Time-series semi-supervised classification. Time series SSC has been studied for many years (Wei and Keogh 2006; Wang et al. 2019). Existing methods based on deep learning for time series SSC mainly use temporal dependencies and time-frequency information for learning. Regarding temporal dependencies, MTL (Jawed, Grabocka, and Schmidt-Thieme 2020) utilizes the sampled subsequence to predict the value of the adjacent next subsequence. Differently, SemiTime (Fan et al. 2021) and SSTSC (Xi et al. 2022) introduce unsupervised temporal relation prediction losses. Regarding time-frequency information, MTFC (Wei et al. 2023) and TS-TFC (Liu et al. 2023b) incorporate time- and frequency-domain features of time series to enable the model to learn the class distribution more effectively. Unlike the aforementioned methods, we use shapelets for time series SSC to improve the interpretability of the model.

Time-series shapelets. Shapelet-based TSC algorithms can be broadly classified into discovery-based and learning-based approaches. The former (Lines et al. 2012; Ji et al. 2019; Li et al. 2020, 2022) typically search for shapelets across all the subsequences within a time series dataset, which is extremely time-consuming. The latter (Grabocka et al. 2014; Ma et al. 2020; Li et al. 2021; Yamaguchi, Ueno, and Kashima 2022) learn shapelets with the help of many labeled time series, which effectively reduces the time to obtain shapelets. However, when labeled data is insufficient, existing learning-based methods rarely consider the use of unlabeled data to improve shapelet quality.

Background

Problem formulation. Suppose a time series data set $\mathcal{D}$, which can be divided into a labeled set $\mathcal{D}^L = \{x|x = (x_i^L, y_i^L)\}$ and an unlabeled set $\mathcal{D}^U = \{x|x = (x_i^U)\}$. Here, $x = \{c_n\}_{n=0}^N$ represents a time series, where N is the sequence length and $c_n \in \mathbb{R}$ is a real value. Additionally, y_i^L corresponds to the target label of the sample x_i^L. It's worth noting that the number of time series samples in $\mathcal{D}^L$ is smaller compared to $\mathcal{D}^U$. As such, we can use $\mathcal{D}^L$ and $\mathcal{D}^U$ for time series semi-supervised classification. Similar to existing SSC methods (Yang et al. 2022), we employ the cross-entropy $\mathcal{L}_{cls}$ for training the model using $\mathcal{D}^L$. In time series SSC, the critical issue is to use $\mathcal{D}^U$ to improve the classification performance of the model.

Diffusion models. There is a forward and a reverse diffusion process in diffusion models (Ho, Jain, and Abbeel 2020; Salimans and Ho 2022; Schneider, Jin, and Schölkopf 2023). The classical forward diffusion process gradually adds Gaussian noise to the original sample x_0 until it becomes a completely random Gaussian distribution x_t. In practice, the noise addition follows a Markovian process, defined as follows:

$$q\left(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}\right) = \mathcal{N}\left(\boldsymbol{x}_t; \sqrt{\alpha_t}\boldsymbol{x}_{t-1}, (1-\alpha_t)\boldsymbol{I}\right), \quad (1)$$

where α_t denotes the noise level added at step t. The above process can be shown:

$$q\left(\boldsymbol{x}_t \mid \boldsymbol{x}_0\right) = \int q\left(\boldsymbol{x}_{1:t} \mid \boldsymbol{x}_0\right) d_{\boldsymbol{x}_{1:t-1}} \sim \mathcal{N}\left(\sqrt{\bar{\alpha}_t}\boldsymbol{x}_0, \left(1-\bar{\alpha}_t\right)\boldsymbol{I}\right), \quad (2)$$

where $\bar{\alpha}_t := \prod_{i=1}^t \alpha_i$. As a result, any $\boldsymbol{x}_t = \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \left(1-\bar{\alpha}_t\right)\epsilon_t$, where $\epsilon \sim \mathcal{N}(0, \boldsymbol{I})$ denotes injected noise. The reverse process involves a learnable neural network $g(\cdot)$ to denoise x_t for recovering x_0, which can be defined as:

$$p\left(\boldsymbol{x}_{0:T}\right) = p\left(\boldsymbol{x}_T\right) \prod_{t=T}^{1} p\left(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t\right), \quad (3)$$

where $p\left(\boldsymbol{x}_T\right) \sim \mathcal{N}(0, \boldsymbol{I})$ is a standard normal distribution. $p\left(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t\right)$ means that x_{t-1} is obtained by removing the

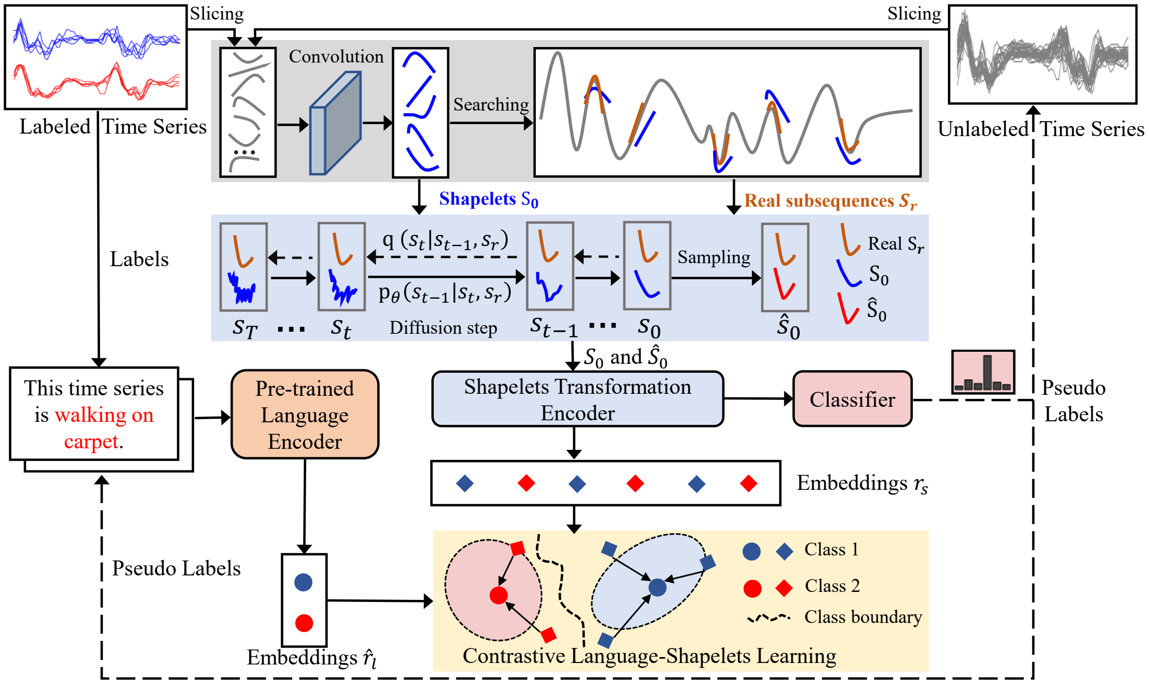


Figure 1: An illustration of the proposed diffusion language-shapelets model. Both learned shapelets $\mathbf{S}_0$ and real subsequences $\mathbf{S}_r$ comprise k shapes, and all of these shapes participate in the diffusion step. To simplify, we depict the learning process for one shape. During training, the classifier's predicted labels serve as pseudo-labels for the unlabeled time series.

estimated Gaussian noise from x_t using $g(\cdot)$. Therefore, the learning objective of the diffusion model can be defined as:

$$\mathcal{L}_\epsilon = \mathbb{E}_{q(\boldsymbol{x}_t|\boldsymbol{x}_0)}\left[\left\|\epsilon_t - g\left(\boldsymbol{x}_t, t\right)\right\|_2^2\right], \tag{4}$$

where ϵ_t denotes the noise to obtain x_t from x_0 in Eq. (2).

The Proposed Method

The Model Overview

The illustration of DiffShape is shown in Figure 1. DiffShape incorporates two mechanisms: (i) self-supervised diffusion learning; (ii) contrastive language-shapelets learning. As for the former, we initially slice all labeled and unlabeled time series with a fixed sliding window to extract real subsequences. These real subsequences are fed into a convolutional layer to obtain the learned shapelets (denoted as $\mathbf{S}_0$). Afterwards, the similarity between $\mathbf{S}_0$ and all real subsequences of each time series is calculated to search for a set of the most similar real subsequences, denoted as $\mathbf{S}_r$. Finally, $\mathbf{S}_0$ and $\mathbf{S}_r$ are fed into a 1-D U-Net (Ronneberger, Fischer, and Brox 2015) network as $g(\cdot)$ for self-supervised learning. It's worth noting that DiffShape uses $\mathbf{S}_r$ as a diffusion condition to guide the model to generate shapelets $\hat{\mathbf{S}}_0$.

In contrastive language-shapelet learning, natural language descriptions are initially generated for time series using labels from labeled samples in $\mathcal{D}^L$ and pseudo-labels from unlabeled samples in $\mathcal{D}^U$. Subsequently, a frozen pre-trained language encoder (Raffel et al. 2020) transforms the generated text descriptions into embeddings $\hat{r}_l$. Meanwhile, a shapelet transformation encoder is employed to convert $\mathbf{S}_0$ and $\hat{\mathbf{S}}_0$ into embeddings r_s. Finally, contrastive learning (Chen et al. 2020) is used to minimize the distance between r_s and $\hat{r}_l$, and r_s is fed to the classifier for training.

Diffusion for Shapelet Generation

To improve the interpretability of the generated shapelets in the absence of labeled data, we design a self-supervised diffusion learning mechanism based on the most similar real subsequences of each time series.

Search for similar real subsequences. We use a sliding window of length L to slice the time series samples in $\mathcal{D}$ to obtain all possible subsequences. Specifically, j is used as the starting point to obtain the subsequence $x_{(i,j+L-1)}$ of sample x_i, which contains the time range $(N_j, ..., N_{(j+L-1)})$. For each time series x in $\mathcal{D}$, we can obtain J subsequences, where $J = N - L + 1$. Thus, the time series x can be denoted as $\mathrm{s} \in \mathbb{R}^{L\times J}$. We input s into a convolutional layer to derive k shapelets s_0^k for classification. Furthermore, we utilize s_0^k of each time series and all the real subsequence $x_{(i,j+L-1)}$ in s to calculate their similarity, which is formulated as follows:

$$M_{i,j} = \max_{j=1,...,J} sim(x_{(i,j+L-1)}, \mathrm{s}_0^i), \tag{5}$$

where $i \in [1, k]$, $sim(\cdot)$ denotes the cosine similarity calculation function, and a larger value implies a greater similarity. We extract the top k values from $M_{i,j}$ to form the real subsequences s_r^k. This search process, performed within each minibatch B using GPUs, has a time complexity of $O(kL(1-\eta)B)$. Hence, obtaining $\mathbf{s}_r^k$ is not time-consuming.

Shapelet diffusion. DiffShape utilizes $\mathbf{S}_0 = \{\mathbf{s}_0^k | \mathbf{s}_0^k = (\mathbf{s}_{0,i}^k)\}$ from labeled and unlabeled time series as the input data for diffusion learning. In DiffShape's forward process, Eq. (2) is used to inject random Gaussian noise into $\mathbf{s}_0^k$ to obtain $\mathbf{s}_t^k$. In the inverse process, Eq. (3) conducts generation training in an unconditional case, which does not guarantee the similarity of $\mathbf{s}_0^k$ to the real subsequences of the time series. To this end, the real subsequence $\mathbf{s}_r^k$ is leveraged as a condition for self-supervised learning, which is as follows:

$$p\left(\mathbf{s}_{0:T}^k \mid \mathbf{s}_r^k\right) = p\left(\mathbf{s}_T^k\right) \prod_{t=T}^{1} p\left(\mathbf{s}_{t-1}^k \mid \mathbf{s}_t^k, \mathbf{s}_r^k\right), \qquad (6)$$

By Eq. (6), we add random noise ϵ and conditional information $p_\theta\left(\mathbf{s}_r^k\right)$ (p_θ denotes $g(\cdot)$) to the shapelets $\mathbf{s}_0^k$ during the reverse process. In this way, we let $p_\theta\left(\mathbf{s}_0^k | \mathbf{s}_r^k\right)$ exploit the distribution of real subsequences from many unlabeled time series for the shapelet generation training. During sampling, the real subsequences $\mathbf{s}_r^k$ are utilized as conditions to guide $g(\cdot)$ to generate new shapelets $\hat{\mathbf{s}}_0^k$. To better utilize $g(\cdot)$ for sampling to obtain $\hat{\mathbf{s}}_0^k$, we introduce a reweighted training strategy based on Eq. (4) for noise estimation:

$$\mathcal{L}_{diff} = E_{q\left(\mathbf{s}_t^k | \mathbf{s}_0^k\right)}\left[\left\| v_{\sigma_t} - g\left(\mathbf{s}_{\sigma_t}^k, \sigma_t, \mathbf{s}_r^k\right)\right\|_2^2\right], \qquad (7)$$

where $v_{\sigma_t} = \alpha_{\sigma_t}\epsilon - \beta_{\sigma_t}\mathbf{s}_0^k$, $\beta_{\sigma_t}^2 = 1 - \alpha_{\sigma_t}^2$, $\alpha_{\sigma_t} = \cos\left(\frac{\pi}{2}\sigma_t\right)$, and $\sigma_t \in [0,1]$. Based on Eq. (7), we incorporate the denoising diffusion implicit model sampler (Song, Meng, and Ermon 2020) for sampling, which achieves a good trade-off between the sampling quality and the number of sampling steps T. The specific sampling process is as follows:

$$\hat{\mathbf{s}}_0^k = \alpha_{\sigma_t}\mathbf{s}_{\sigma_t}^k - \beta_{\sigma_t} g\left(\mathbf{s}_{\sigma_t}^k, \sigma_t, \mathbf{s}_r^k\right), \qquad (8)$$

$$\hat{\mathbf{s}}_{\sigma_{t-1}}^k = \alpha_{\sigma_{t-1}}\hat{\mathbf{s}}_0^k + \beta_{\sigma_{t-1}}\left(\beta_{\sigma_t}\mathbf{s}_{\sigma_t}^k + \alpha_{\sigma_t} g\left(\mathbf{s}_{\sigma_t}^k, \sigma_t, \mathbf{s}_r^k\right)\right), \qquad (9)$$

where $\mathbf{s}_{\sigma_t}^k \sim \mathcal{N}(0, \mathrm{I})$ at the first iteration. During each iteration, we utilize $\hat{\mathbf{s}}_{\sigma_{t-1}}^k$ as $\mathbf{s}_{\sigma_t}^k$ in Eq. (8) until $t = 0$. Note that the $\mathbf{s}_r^k$ of the labeled time series is solely used as the condition to generate $\hat{\mathbf{s}}_0^k$, and let $\hat{\mathbf{S}}_0 = \{\hat{\mathbf{s}}_0^k | \hat{\mathbf{s}}_0^k = (\hat{\mathbf{s}}_{0,i}^k)\}$. This strategy provides two primary advantages. First, it reduces the runtime in the sampling process due to the number of labeled samples being small. Simultaneously, it employs the gradient information from the classifier to guide the training of shapelet generation. Second, it helps to reduce the classification errors caused by the generated shapelets $\hat{\mathbf{s}}_0^k$ because a large number of unlabeled samples lack labels.

Contrastive Language-Shapelets Learning

This subsection discusses how a contrastive language-shapelet learning mechanism enhances the discriminative power of the generated shapelets.

Natural language construction. The label information of the time series is utilized to create the natural language description of each sample. Concretely, we initially formulate a text template, also referred to as a named hard prompt (Liu et al. 2023a; Khattak et al. 2023). For example, we can use "This time series is ____." as a hard prompt. Then, we impute the blanks "____" in the prompt based on the keyword information associated with the classes of the time series dataset. As shown in Figure 1, we construct a natural language description of the *SonyAIBORobotSurface1* UCR (Dau et al. 2019) time series dataset. Based on the information provided by the dataset provider (Vail and Veloso 2004), the *SonyAIBORobotSurface1* dataset contains two classes: walking on carpet and cement. Therefore, natural language descriptions are constructed using the labels of labeled data and pseudo-labels of unlabeled data. To reduce the classification errors caused by incorrect labels in the pseudo-labels, we choose the predicted soft labels with high confidence of the classifier (Lee et al. 2013; Zhang et al. 2021) as pseudo-labels.

Language-shapelets training. In recent years, contrastive learning (Chen et al. 2020) has performed excellently in time series representation learning (Ma et al. 2023). Specifically, contrastive learning trains a model by decreasing the distance between pairs of positive samples and increasing the distance between pairs of positive and negative samples. In this study, we utilize the labels of time series to construct language-shapelets pairs for contrastive learning, so as to improve the discriminability of the generated shapelets by exploiting the rich semantics of natural language descriptions about time series. To achieve this, a shapelet transformation encoder is employed to transform the shapelets $\mathbf{s}_0^k$ and $\hat{\mathbf{s}}_0^k$ into embeddings r_s. Meanwhile, a frozen pre-trained T5 language encoder (Raffel et al. 2020) converts the natural language descriptions into embeddings r_l. In particular, we use a projection head $h(\cdot)$ consisting of a two-layer nonlinear network that enables the dimension of r_l to be consistent with r_s, denoted as $\hat{r}_l = h(r_t)$. Thus, the training objective for contrastive language-shapelets learning is defined as:

$$\mathcal{L}_{lan} = \frac{-1}{B}\sum_{i=1}^{B}\sum_{j=1}^{B}\mathbf{1}_{y_{ij}=1}\log\frac{\exp\left(\operatorname{sim}\left(r_{s,i}, \hat{r}_{l,j}\right)/\tau\right)}{\sum_{c=1}^{B}\mathbf{1}_{y_{ic}\neq 1}\exp(\operatorname{sim}\left(r_{s,i}, \hat{r}_{l,c}\right)/\tau)}, \qquad (10)$$

where B denotes the number of samples, and τ is a temperature parameter that controls the contrastive learning process. $\mathbf{1}_{y_{ij}=1}$ means that the value is 1 when $r_{s,i}$ and $\hat{r}_{l,j}$ belong to the same class; Otherwise, the value is 0.

The Overall Training Process

As shown in Figure 1, DiffShape utilizes r_s and $\hat{r}_l$ for contrastive language-shapelets learning using Eq. (10). On the other hand, r_s is fed into a classifier using $\mathcal{L}_{cls}$ for classification training. For pseudo-labeled samples, the pseudo-label is used as the ground truth label for training. Practically, the classifier consists of one layer of a linear neural network. Thus, the overall training objective is as follows:

$$\mathcal{L}_{total} = \mathcal{L}_{cls} + \mu_{\text{diff}}\,\mathcal{L}_{\text{diff}} + \mu_{\text{lan}}\,\mathcal{L}_{\text{lan}}, \qquad (11)$$

where the values of μ_{diff} and μ_{lan} are in the range of [0,1], which are hyperparameters used to adjust the training loss ratio. To increase the diversity of shapelets throughout the model training process, we incorporate a $l_{\text{reg}} = \sum_{i\neq j}^{k}\exp\left(\operatorname{sim}\left(\mathbf{s}_0^i, \mathbf{s}_0^j\right)\right)$ as a regularization term to $\mathcal{L}_{total}$, so as to increase the difference between different shapelets. In addition, the pseudo-code for DiffShape is presented in Algorithm 1 within the Appendix.

Labeling Ratio	10%			20%			40%		
Method	Avg. Rank	Win	P-value	Avg. Rank	Win	P-value	Avg. Rank	Win	P-value
Supervised* (Cross entropy)	5.80	4	2.72E-06	5.54	3	3.52E-06	5.50	7	3.01E-05
Pseudo-Label* (Lee et al. 2013)	5.14	4	1.38E-05	4.93	6	1.23E-05	5.22	7	1.26E-05
TE* (Laine and Aila 2016)	4.92	5	4.14E-05	5.04	6	3.68E-05	5.21	6	3.18E-05
LPDeepSSL* (Iscen et al. 2019)	6.15	4	4.26E-07	6.92	3	2.57E-08	6.32	5	7.130E-06
MTL* (Jawed et al. 2020)	8.57	1	8.46E-18	8.62	3	3.10E-17	8.71	4	2.36E-16
TS-TCC* (Eldele et al. 2021)	10.50	0	7.41E-26	10.36	0	3.56E-24	10.29	0	3.95E-24
SemiTime* (Fan et al. 2021)	5.00	8	4.77E-05	4.46	10	4.82E-05	4.37	11	3.69E-04
SSSTC (Xi et al. 2022)	3.92	19	1.65E-04	3.98	17	2.06E-05	3.73	18	6.39E-04
MTFC (Wei et al. 2023)	8.91	3	4.24E-21	9.01	2	1.19E-21	9.26	2	1.25E-19
TS-TFC* (Liu et al. 2023b)	3.24	25	1.51E-02	3.00	27	2.05E-02	2.88	28	3.85E-02
DiffShape (Ours)	**2.92**	**51**	-	**2.86**	**58**	-	**2.62**	**64**	-

Table 1: Test classification accuracy comparisons on 106 UCR time series datasets. * denotes that the test classification accuracies of the baseline are collected from TS-TFC (Liu et al. 2023b). Win denotes the number of datasets in which the corresponding baseline achieved the best test accuracy. The best is in **bold**.

Experiments

Datasets. We used the UCR time series archive (Dau et al. 2019) to evaluate the proposed method. Similar to prior time series SSC work (Liu et al. 2023b), we selected 106 UCR time series datasets for our experiments. Following the suggestion given by Dau et al. (2019); Liu et al. (2023b), we adopted a five-fold cross-validation method, where the training-validation-test set ratio is set to 60%-20%-20% for each dataset. We also randomly selected 10%, 20%, and 40% of the samples in the training set as labeled data, and used the rest as unlabeled data. Additional details regarding the 106 UCR datasets are available in Appendix A.

Baselines. DiffShape is compared with 10 SSC methods, including Supervised, Pseudo-Label (Lee et al. 2013), Temporal Ensembling (TE) (Laine and Aila 2016), LPDeepSSL (Iscen et al. 2019), MTL (Jawed, Grabocka, and Schmidt-Thieme 2020), TS-TCC (Eldele et al. 2021), SemiTime (Fan et al. 2021), SSSTC (Xi et al. 2022), MTFC (Wei et al. 2023), TS-TFC (Liu et al. 2023b). Supervised methods only use labeled data for classification training via cross-entropy. Additionally, we select 4 shapelet-based TSC methods for time series SSC analysis, including Shapelet Transform (ST) (Lines et al. 2012), Learning Time-series Shapelets (LTS) (Grabocka et al. 2014), Fast Shapelet Selection (FSS) (Ji et al. 2019), and Adversarial Dynamic Shapelet Networks (ADSN) (Ma et al. 2020). For more details about baselines, please refer to Appendix B.

Parameter settings. The maximum epoch, the learning rate and the batch size are set to 1000, 1e-3 and 128, respectively. We set μ_{diff} to 0.01, μ_{lan} to 0.001, sampling steps T to 10, and τ in Eq. (10) to 50. Like Liu et al. (2023b), we also use labeled data for warm-up training in the first 300 epochs. Semi-supervised classification aims to enhance the performance of the same architectural model (or encoder) by using unlabeled data (Oliver et al. 2018). Accordingly, the FCN model (Wang, Yan, and Oates 2017) is used as an encoder to obtain shapelet transformation embeddings, and the baselines use the same encoder for fair comparisons. The number of shapelets is $k \in \{2, 5, 10\}$. The length of the shapelet is set to be the η ratio of the time series length, where $\eta \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}$. Like Grabocka et al. (2014), we use a cross-validation grid search method to select k and η. All experiments are conducted with five random seeds, and the averaged test accuracies are reported. We run experiments using PyTorch 1.10 on two NVIDIA GeForce RTX 3090 GPUs. The implementation of DiffShape, along with the supplementary materials provided in the Appendix, can be accessed at https://github.com/qianlima-lab/DiffShape.

Method	Avg. Rank	Win	P-value
ST (Lines et al. 2012)	2.92	0	1.60E-03
LTS (Grabocka et al. 2014)	3.50	0	3.78E-04
FSS (Ji et al. 2019)	3.92	0	2.86E-05
ADSN (Ma et al. 2020)	3.58	0	1.44E-03
DiffShape (Ours)	**1.00**	**12**	-

Table 2: Test classification accuracy comparisons on 12 UCR time series datasets with a 10% labeling ratio.

Main Results

As shown in Table 1, it is found that DiffShape achieves the best classification performance under different labeling ratios on the 106 UCR time series datasets. Among the baseline methods, both MTL and MTFC employ unsupervised time prediction loss for learning unlabeled data, yet fail to enhance the model's classification performance. SemiTime and SSSTC utilize temporal prediction loss as a consistency regularization strategy, proving effective in the context of time series SSC. Compared with the supervised method, Pseudo-Label and TS-TFC use pseudo-labeling techniques that can effectively alleviate the problem of lacking labeled data. In addition, we apply the Wilcoxon signed rank test (Demšar 2006) to assess the significance of test classification accuracies. The results reveal that DiffShape's classification performance is significantly superior (P-value $<$ 0.05) to that of all the considered baselines. For additional insights, a critical difference diagram and the detailed results of Table 1 are provided in Appendix C.

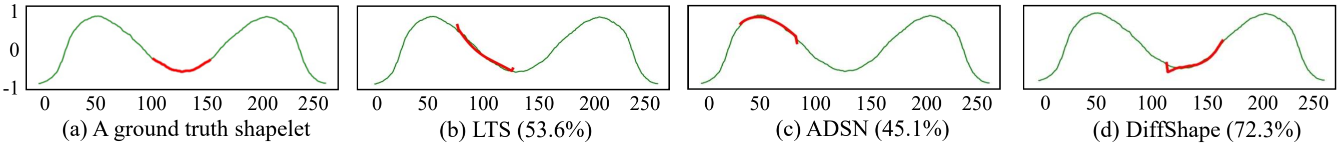


Figure 2: The visualization of shapelet on the *ArrayHead* dataset with a 10% labeling ratio. The test accuracy is in parentheses. (a) represents a ground truth shapelet of Clovis arrowhead class. The position of shapelet learned by (b) LTS and (c) ADSN is away from the ground truth in (a), while the position of shapelet obtained by (d) DiffShape is closer to the ground truth in (a).

	2 labels			5 labels			10 labels		
Method	Avg. Rank	Win	P-value	Avg. Rank	Win	P-value	Avg. Rank	Win	P-value
Supervised (Cross entropy)	4.25	0	5.28E-03	4.50	0	3.07E-03	4.58	0	3.95E-05
LTS (Grabocka et al. 2014)	7.67	0	8.07E-07	7.83	0	1.37E-07	7.83	0	1.28E-07
ADSN (Ma et al. 2020)	7.08	0	6.81E-05	7.00	0	1.44E-05	7.08	0	5.09E-06
SemiTime (Fan et al. 2021)	3.50	0	2.14E-02	3.25	0	3.96E-02	3.17	1	1.94E-02
SSSTC (Xi et al. 2022)	2.92	0	2.18E-04	5.83	0	8.12E-04	5.33	1	2.51E-03
MTFC (Wei et al. 2023)	5.92	0	2.08E-02	3.00	0	1.97E-02	2.83	2	3.93E-02
TS-TFC (Liu et al. 2023b)	3.08	3	8.01E-03	2.58	3	3.38E-02	2.75	3	3.56E-02
DiffShape (Ours)	**1.42**	**10**	-	**1.67**	**9**	-	**1.75**	**8**	-

Table 3: Test classification accuracy on 12 UCR time series datasets with few labels per class without using unlabeled data.

Comparisons with Shapelet-based TSC Methods

To analyze the performance of shapelets obtained by DiffShape on classification, we perform a comparative experimental analysis using shapelet-based TSC methods combined with the pseudo-labeling technique in DiffShape for SSC. Specifically, LTS (Grabocka et al. 2014) and ADSN (Ma et al. 2020) disclose shapelet learning hyperparameters for 28 and 18 UCR datasets, respectively. To reduce the negative impact of UCR datasets with small sample sizes on the classification stability, we selected 12 UCR datasets shared by both LTS and ADSN for experimental analysis. Table 2 shows the statistical classification results on the 12 UCR time series datasets with a 10% labeling ratio. The detailed results of Table 2 and the results with labeling ratios of 20% and 40% are provided in Appendix D. Compared to ST, LTS, FSS and ADSN, it is found that the shapelets obtained by DiffShape are more favorable for time series SSC.

Results on a Few Labeled Time Series

To verify the efficacy of DiffShape in mitigating the issue of lacking labeled samples, we perform classification analyses on time series datasets with only a few labels per class without using unlabeled data. Specifically, we select Supervised, LTS (Grabocka et al. 2014), ADSN (Ma et al. 2020), SemiTime (Fan et al. 2021), SSSTC (Xi et al. 2022), MTFC (Wei et al. 2023) and TS-TFC (Liu et al. 2023b) as baselines. Similar to the previous section, we employ the 12 UCR time series datasets with only 2, 5, and 10 labeled samples per class for analyses. As shown in Table 3, we find that the Avg. Rank and Win metrics achieved by DiffShape are better than those of 5 and 10 labeled samples when only 2 labeled samples are available, and both of them are better than those of baselines. The above results demonstrate that DiffShape can alleviate the lack of labeled time series data. For detailed results of Table 3, please refer to Appendix E.

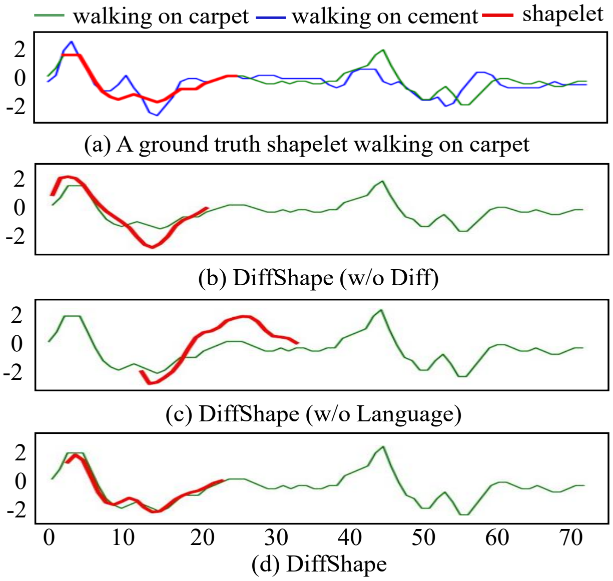
Ablation Analysis

To assess the individual effectiveness of each module within DiffShape, we choose the 12 UCR time series with a 10% labeling ratio like Table 3 for the experiments. The statistical ablation results are reported in Table 4. For detailed results, please refer to Appendix F. Specifically, (1) **w/o Diff**: we remove the self-supervised diffusion learning mechanism from DiffShape; (2) **real subsequence**: For $\hat{\mathbf{S}}_0$ generated by DiffShape, we use the most similar real subsequence to replace $\hat{\mathbf{S}}_0$ for training; (3) **random shape**: we use a randomly selected subsequence (or shape) from the real subsequence of each time series as a condition for self-supervised diffusion learning; (4) **w/o Language**: we remove the contrastive language-shapelets learning mechanism from DiffShape; (5) **w/o Diff** & **Language**: we remove the self-supervised diffusion learning and contrastive language-shapelets learning mechanisms from DiffShape.

As shown in Table 4, both the self-supervised diffusion learning mechanism and the contrastive Language-Shapelets learning mechanism are able to effectively improve the classification performance of DiffShape. In particular, *real subsequence* and *random shape* results show that the self-supervised diffusion mechanism could utilize the distribution of a large number of unlabeled samples to enable the generated $\hat{\mathbf{S}}_0$ to be more conducive for improving classification performance, thus alleviating the issue of lacking labeled time series samples. In addition, a runtime analysis of DiffShape is presented in Appendix G.

Visualization Analysis

In this subsection, we analyze the interpretability of shapelets generated by DiffShape. The *ArrowHead* dataset in the UCR archive aims to classify the shapes of the projectile points of a notch in an arrow, which contains three



(a) A ground truth shapelet walking on carpet

(b) DiffShape (w/o Diff)

(c) DiffShape (w/o Language)

(d) DiffShape

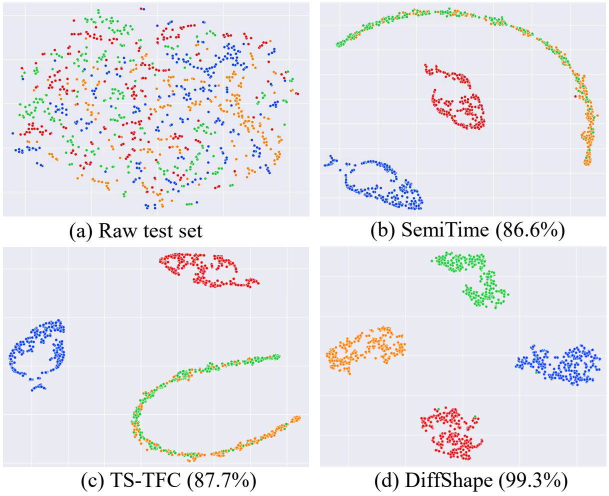
Figure 3: One sample from the *SonyAIBORobotSurface1* dataset with a 10% labeling ratio and the obtained shapelet with the smallest distance between the real shape.

Method	Avg. Rank	Win	P-value
DiffShape	**1.00**	**12**	-
w/o Diff	3.08	1	3.35E-02
real subsequence	2.83	1	4.39E-02
random shape	3.17	1	4.14E-02
w/o Language	3.67	1	3.04E-02
w/o Diff & Language	5.17	0	1.22E-02

Table 4: Ablation study results of DiffShape on 12 UCR time series datasets with a 10% labeling ratio.

types of arrow heads, i.e., *Avonlea*, *Clovis*, and *Mix*. Ye and Keogh (2009) demonstrated that the shape of *Clovis* at the [100,150] segment can be indicative of the *Clovis* class (see Figure 2 (a)). We select LTS (Grabocka et al. 2014), ADSN (Ma et al. 2020), and DiffShape for SSC on the *ArrowHead* dataset. We then visualize the learned best shapelet for a *Clovis* arrowhead sample from the test set in Figure 2. Compared to LTS and ADSN, it is found that DiffShape generates a more discriminative shape, contributing to better classification performance as well as interpretability.

We also choose the *SonyAIBORobotSurface1* dataset to investigate the role of different components of DiffShape in the shapelet generation. The *SonyAIBORobotSurface1* dataset involves two distinct actions: walking on cement and carpet. Notably, Mueen, Keogh, and Young (2011) indicate that subsequences within the interval of [2,23] are identified as the most discriminatory shapelet (see Figure 3 (a)). Figure 3 (b) and (d) show that **w/o Diff** in DiffShape could result in the obtained shapelet differing significantly from the original subsequence. Comparing Figure 3 (c) and (d), **w/o Language** in DiffShape could lead to the difference between the obtained shapelet and the best ground truth shapelet position. The ablation results of **w/o Language** in Table 4 show that contrastive language-shapelets learning improves the classification performance of shapelets. In other words, **w/o Language** in DiffiShape causes the generated shapelets to deviate from the position of the best ground truth shapelet. Please refer to Figure 2 in the Appendix for comparing DiffShape with the shapelet learned by LTS and ADSN on the *SonyAIBORobotSurface1* dataset.



(a) Raw test set

(b) SemiTime (86.6%)

(c) TS-TFC (87.7%)

(d) DiffShape (99.3%)

Figure 4: The t-SNE visualization on the *TwoPatterns* dataset with a 10% labeling ratio. The test accuracy is in parentheses.

In addition, we employ the t-SNE (Van der Maaten and Hinton 2008) technique to analyze the embeddings learned by SemiTime, TS-TFC, and DiffShape. As shown in Figure 4 (a), the original time series test set of *TwoPatterns* exhibits mixed sample classes. While the embeddings learned by SemiTime and TF-TFC distinguish class 0 (blue dots) and class 3 (red dots), they struggle to differentiate class 1 (orange dots) and class 2 (green dots). In contrast, Figure 4 (d) demonstrates that DiffShape can clearly distinguish the four classes of TwoPatterns, highlighting the more discriminative nature of shapelets obtained through DiffShape. To further validate the effectiveness of DiffShape, we present the t-SNE visualization for the *UWaveGestureLibraryAll* time series dataset in Figure 3 of the Appendix.

Conclusion

In this paper, we propose a diffusion Language-Shapelets model for semi-supervised classification of time series. In particular, a self-supervised diffusion learning mechanism is designed to induce the generated shapelets to become more similar to the real subsequences. We further introduce a contrastive language-shaplets learning mechanism to encourage the generated shapelets to be more discriminative. Extensive experiments on the UCR time series archive proved that the proposed DiffShape method has advanced classification performance and good interpretability. In the future, we aim to explore multivariate time series shapelet models for SSC.

Acknowledgments

We thank the anonymous reviewers for their helpful feedbacks. We thank Professor Eamonn Keogh and all the people who have contributed to the UCR time series classification archive. The work described in this paper was partially funded by the National Natural Science Foundation of China (Grant Nos. 62272173, 62206041,61872148), the Natural Science Foundation of Guangdong Province (Grant Nos. 2022A1515010179, 2019A1515010768), the Science and Technology Planning Project of Guangdong Province (Grant No. 2023A0505050106), and China University Industry-University-Research Innovation Fund under grants 2022IT174. The authors would like to thank Junhao Zheng from SCUT for the technical discussions.

References

Chen, K.; Zhang, D.; Yao, L.; Guo, B.; Yu, Z.; and Liu, Y. 2021. Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities. *ACM Computing Surveys (CSUR)*, 54(4): 1–40.

Chen, T.; Kornblith, S.; Norouzi, M.; and Hinton, G. 2020. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, 1597–1607. PMLR.

Cheng, Z.; Yang, Y.; Wang, W.; Hu, W.; Zhuang, Y.; and Song, G. 2020. Time2graph: Revisiting time series modeling with dynamic shapelets. In *Proceedings of the AAAI conference on artificial intelligence*, 3617–3624.

Dau, H. A.; Bagnall, A.; Kamgar, K.; Yeh, C.-C. M.; Zhu, Y.; Gharghabi, S.; Ratanamahatana, C. A.; and Keogh, E. 2019. The UCR time series archive. *IEEE/CAA Journal of Automatica Sinica*, 6(6): 1293–1305.

Demšar, J. 2006. Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine learning research*, 7: 1–30.

Eldele, E.; Ragab, M.; Chen, Z.; Wu, M.; Kwoh, C. K.; Li, X.; and Guan, C. 2021. Time-series representation learning via temporal and contextual contrasting. *arXiv preprint arXiv:2106.14112*.

Fan, H.; Zhang, F.; Wang, R.; Huang, X.; and Li, Z. 2021. Semi-supervised time series classification by temporal relation prediction. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 3545–3549. IEEE.

Grabocka, J.; Schilling, N.; Wistuba, M.; and Schmidt-Thieme, L. 2014. Learning time-series shapelets. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 392–401.

Gruver, N.; Finzi, M. A.; Qiu, S.; and Wilson, A. G. 2023. Large Language Models Are Zero-Shot Time Series Forecasters. In *Thirty-seventh Conference on Neural Information Processing Systems*.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33: 6840–6851.

Iscen, A.; Tolias, G.; Avrithis, Y.; and Chum, O. 2019. Label propagation for deep semi-supervised learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5070–5079.

Jawed, S.; Grabocka, J.; and Schmidt-Thieme, L. 2020. Self-supervised learning for semi-supervised time series classification. In *Advances in Knowledge Discovery and Data Mining: 24th Pacific-Asia Conference, PAKDD 2020, Singapore, May 11–14, 2020, Proceedings, Part I 24*, 499–511. Springer.

Ji, C.; Zhao, C.; Liu, S.; Yang, C.; Pan, L.; Wu, L.; and Meng, X. 2019. A fast shapelet selection algorithm for time series classification. *Computer networks*, 148: 231–240.

Jin, M.; Wen, Q.; Liang, Y.; Zhang, C.; Xue, S.; Wang, X.; Zhang, J.; Wang, Y.; Chen, H.; Li, X.; et al. 2023. Large models for time series and spatio-temporal data: A survey and outlook. *arXiv preprint arXiv:2310.10196*.

Khattak, M. U.; Rasheed, H.; Maaz, M.; Khan, S.; and Khan, F. S. 2023. Maple: Multi-modal prompt learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 19113–19122.

Laine, S.; and Aila, T. 2016. Temporal ensembling for semi-supervised learning. *arXiv preprint arXiv:1610.02242*.

Lee, D.-H.; et al. 2013. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML*, 896. Atlanta.

Li, G.; Choi, B.; Xu, J.; Bhowmick, S. S.; Chun, K.-P.; and Wong, G. L.-H. 2020. Efficient shapelet discovery for time series classification. *IEEE transactions on knowledge and data engineering*, 34(3): 1149–1163.

Li, G.; Choi, B.; Xu, J.; Bhowmick, S. S.; Chun, K.-P.; and Wong, G. L.-H. 2021. Shapenet: A shapelet-neural network approach for multivariate time series classification. In *Proceedings of the AAAI conference on artificial intelligence*, 8375–8383.

Li, G.; Choi, B.; Xu, J.; Bhowmick, S. S.; Mah, D. N.-y.; and Wong, G. L. 2022. IPS: Instance Profile for Shapelet Discovery for Time Series Classification. In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, 1781–1793. IEEE.

Lin, L.; Xu, B.; Wu, W.; Richardson, T. W.; and Bernal, E. A. 2019. Medical Time Series Classification with Hierarchical Attention-based Temporal Convolutional Networks: A Case Study of Myotonic Dystrophy Diagnosis. In *CVPR workshops*, 83–86.

Lines, J.; Davis, L. M.; Hills, J.; and Bagnall, A. 2012. A shapelet transform for time series classification. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 289–297.

Liu, P.; Yuan, W.; Fu, J.; Jiang, Z.; Hayashi, H.; and Neubig, G. 2023a. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9): 1–35.

Liu, Z.; Ma, Q.; Ma, P.; and Wang, L. 2023b. Temporal-Frequency Co-training for Time Series Semi-supervised

Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 8923–8931.

Ma, Q.; Liu, Z.; Zheng, Z.; Huang, Z.; Zhu, S.; Yu, Z.; and Kwok, J. T. 2023. A Survey on Time-Series Pre-Trained Models. *arXiv preprint arXiv:2305.10716*.

Ma, Q.; Zhuang, W.; Li, S.; Huang, D.; and Cottrell, G. 2020. Adversarial dynamic shapelet networks. In *Proceedings of the AAAI conference on artificial intelligence*, 5069–5076.

Maweu, B. M.; Dakshit, S.; Shamsuddin, R.; and Prabhakaran, B. 2021. CEFEs: a CNN explainable framework for ECG signals. *Artificial Intelligence in Medicine*, 115: 102059.

Mueen, A.; Keogh, E.; and Young, N. 2011. Logical-shapelets: an expressive primitive for time series classification. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1154–1162.

Oliver, A.; Odena, A.; Raffel, C. A.; Cubuk, E. D.; and Goodfellow, I. 2018. Realistic evaluation of deep semi-supervised learning algorithms. *Advances in neural information processing systems*, 31.

Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, 8748–8763. PMLR.

Raffel, C.; Shazeer, N.; Roberts, A.; Lee, K.; Narang, S.; Matena, M.; Zhou, Y.; Li, W.; and Liu, P. J. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research*, 21(1): 5485–5551.

Rasul, K.; Seward, C.; Schuster, I.; and Vollgraf, R. 2021. Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In *International Conference on Machine Learning*, 8857–8868. PMLR.

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

Salimans, T.; and Ho, J. 2022. Progressive distillation for fast sampling of diffusion models. *arXiv preprint arXiv:2202.00512*.

Schneider, F.; Jin, Z.; and Schölkopf, B. 2023. Mo\^ usai: Text-to-Music Generation with Long-Context Latent Diffusion. *arXiv preprint arXiv:2301.11757*.

Shen, L.; and Kwok, J. 2023. Non-autoregressive Conditional Diffusion Models for Time Series Prediction. *International Conference on Machine Learning*.

Song, J.; Meng, C.; and Ermon, S. 2020. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*.

Tashiro, Y.; Song, J.; Song, Y.; and Ermon, S. 2021. Csdi: Conditional score-based diffusion models for probabilistic time series imputation. *Advances in Neural Information Processing Systems*, 34: 24804–24816.

Vail, D.; and Veloso, M. 2004. Learning from accelerometer data on a legged robot. *IFAC Proceedings Volumes*, 37(8): 822–827.

Van der Maaten, L.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of machine learning research*, 9(11).

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

Wang, Z.; Yan, W.; and Oates, T. 2017. Time series classification from scratch with deep neural networks: A strong baseline. In *2017 International joint conference on neural networks (IJCNN)*, 1578–1585. IEEE.

Wei, C.; Wang, Z.; Yuan, J.; Li, C.; and Chen, S. 2023. Time-frequency based multi-task learning for semi-supervised time series classification. *Information Sciences*, 619: 762–780.

Wei, L.; and Keogh, E. 2006. Semi-supervised time series classification. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 748–753.

Xi, L.; Yun, Z.; Liu, H.; Wang, R.; Huang, X.; and Fan, H. 2022. Semi-supervised time series classification model with self-supervised learning. *Engineering Applications of Artificial Intelligence*, 116: 105331.

Yamaguchi, A.; Ueno, K.; and Kashima, H. 2022. Learning time-series shapelets enhancing discriminability. In *Proceedings of the 2022 SIAM International Conference on Data Mining (SDM)*, 190–198. SIAM.

Yang, X.; Song, Z.; King, I.; and Xu, Z. 2022. A survey on deep semi-supervised learning. *IEEE Transactions on Knowledge and Data Engineering*.

Ye, L.; and Keogh, E. 2009. Time series shapelets: a new primitive for data mining. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 947–956.

Yuan, X.; Li, L.; Shardt, Y. A.; Wang, Y.; and Yang, C. 2020. Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development. *IEEE Transactions on Industrial Electronics*, 68(5): 4404–4414.

Zhang, B.; Wang, Y.; Hou, W.; Wu, H.; Wang, J.; Okumura, M.; and Shinozaki, T. 2021. Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling. *Advances in Neural Information Processing Systems*, 34: 18408–18419.

Zhang, X.; Chowdhury, R. R.; Zhang, J.; Hong, D.; Gupta, R. K.; and Shang, J. 2023. Unleashing the Power of Shared Label Structures for Human Activity Recognition. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, 3340–3350.