Appendix:
A Unified Shape-Aware Foundation Model for Time Series Classification

Zhen Liu[1,2], Yucheng Wang[2], Boyuan Li[1], Junhao Zheng[1], Emadeldeen Eldele[3], Min Wu[2*], Qianli Ma[1*]

[1]School of Computer Science and Engineering, South China University of Technology, Guangzhou, China
[2]Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore
[3]Department of Computer Science, Khalifa University, UAE
cszhenliu@foxmail.com, wumin@a-star.edu.sg, qianlima@scut.edu.cn

A Details of Experimental Setting

A.1 Dataset Description

The large-scale source dataset is constructed by aggregating training samples from the UCR archive (Dau et al. 2019), the UEA archive (Bagnall et al. 2018), and eight time series datasets across diverse domains (Eldele et al. 2021; Zhang et al. 2022), including: Epilepsy (Andrzejak et al. 2001), SleepEEG (Kemp et al. 2000), HAR (Anguita et al. 2013), Gesture (Liu et al. 2009), FD-A (Lessmeier et al. 2016), FD-B (Lessmeier et al. 2016), ECG (Clifford et al. 2017), and EMG (Goldberger et al. 2000). For the aforementioned 166 datasets, most time series exhibit varying sequence lengths, making them unsuitable as direct inputs for deep learning models. To address this inconsistency, following the approach of (Lin et al. 2024), we apply an interpolation algorithm to uniformly rescale all sequences to a fixed length of 512 for model pre-training. It is worth noting that 30 datasets from the UEA archive and 8 additional datasets contain multivariate time series, with varying numbers of variables across datasets. To handle this inconsistency, following the method of (Lin et al. 2024; Feofanov et al. 2025), we adopt a channel-independent strategy that converts each multivariate time series into multiple univariate sequences corresponding to its channels, which are then used as training samples in the source dataset.

To capture multi-scale temporal patterns during pretraining, five sliding window configurations with lengths and strides $W_q = K_q \in \{64, 32, 16, 8, 4\}$ are applied, yielding {8, 16, 32, 64, 128} non-overlapping subsequences per sample, respectively. The choice of sliding window size is based on two main considerations. First, since all pre-training sequence samples are standardized to a length of 512, setting the maximum non-overlapping window size to 64 provides a relatively large subsequence configuration. Second, our preliminary experiments show that extending the maximum window length to 128 or 256, which correspond to 4 and 2 non-overlapping subsequences (shape tokens), respectively, yields limited improvement in fine-tuning classification performance while significantly increasing pre-training time and GPU memory consumption. Therefore, the maximum sliding window size during pre-training is set to 64. Additionally, during pre-training, each sequence is augmented twice through random cropping and encoded within the MoCo v3 contrastive learning framework (Chen, Xie, and He 2021), which employs momentum encoders to generate additional views, resulting in 992 shape tokens per sample. In total, the pre-training time series source dataset contains approximately 1.9 billion shape tokens derived from 1.89 million samples.

During fine-tuning, the sliding window size (W_q) follows the same configuration as in the pre-training stage, while the stride size (K_q) is uniformly set to 4, or 2 for sequences shorter than 64, to increase the number of subsequences used for classification. This design helps preserve discriminative subsequence information and mitigates potential loss during training. It is noteworthy that most sequences in the 128 UCR datasets for fine-tuning and the 30 additional datasets for zero-shot evaluation have lengths shorter than 600, making a sliding window size of 64 a reasonable choice. For datasets with sequence lengths shorter than 64, we exclude cases where the window size exceeds the sequence length and adopt smaller window sizes instead. In particular, the smallest window size of 4 ensures successful fine-tuning across all datasets.

For downstream classification evaluation, we employ a total of 158 univariate time series classification datasets. Among them, 128 datasets are sourced from the UCR archive (Dau et al. 2019), widely recognized in the time series classification field for their coverage of real-world domains such as healthcare, motion capture, and sensor data. The datasets are publicly available at https://www.cs.ucr.edu/~eamonn/time_series_data_2018/. Following the protocol in (Ma et al. 2024), datasets containing missing values are imputed using the mean of available observations at each timestamp across the training set. Detailed statistics for each UCR dataset are summarized in Table 1.

To assess the cross-domain generalization ability of FMs, we further evaluate *UniShape* in a zero-shot setting using 30 additional univariate datasets from various sources, none of which are used during pre-training. These datasets are publicly released by the authors of the time series classification benchmark (Middlehurst, Schäfer, and Bagnall 2024), available at https://tsml-eval.readthedocs.io/en/latest/publications/2023/tsc_bakeoff/tsc_bakeoff_2023.html. Fol-

*Corresponding authors.
Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

lowing their preprocessing protocol (Middlehurst, Schäfer, and Bagnall 2024), missing values are handled either by discarding incomplete instances or by imputing with the sequence mean perturbed by small Gaussian noise. For consistency, we adopt the processed 30 datasets by (Middlehurst, Schäfer, and Bagnall 2024) for our zero-shot classification analysis. A summary of dataset statistics for these 30 datasets is presented in Table 2.

A.2 Baseline Description

We evaluate UniShape against 16 baseline methods, categorized into three groups:

- **Non-Deep Learning (NDL)**: Rocket (Dempster, Petitjean, and Webb 2020), MiniRocket (Dempster, Schmidt, and Webb 2021), RDST (Guillaume, Vrain, and Elloumi 2022), MultiRocket-Hydra (MR-H) (Dempster, Schmidt, and Webb 2023);
- **Domain-Specific Deep Learning (DS)**: InceptionTime (Ismail Fawaz et al. 2020), TS2Vec (Yue et al. 2022), PatchTST (Nie et al. 2023), TimesNet (Wu et al. 2023), SoftShape (Liu et al. 2025);
- **Foundation Models (FM)**: GPT4TS (Zhou et al. 2023), MOMENT (Goswami et al. 2024), UniTS (Gao et al. 2024), NuTime (Lin et al. 2024), Mantis (Feofanov et al. 2025). Also, we include two forecasting-based FMs for zero-shot classification: Chronos (Ansari et al. 2024) and Moirai (Woo et al. 2024).

We briefly describe the detailed information of each method below:

- **Rocket** (Dempster, Petitjean, and Webb 2020): Employs thousands of random convolutional kernels as fixed feature extractors followed by a linear classifier for time series classification.
  Source code: https://github.com/aeon-toolkit/aeon/blob/main/aeon/classification/convolution_based/_rocket.py
- **MiniRocket** (Dempster, Schmidt, and Webb 2021): Improved Rocket by introducing deterministic kernels and reduced operations, significantly improving speed while maintaining accuracy.
  Source code: https://github.com/aeon-toolkit/aeon/blob/main/aeon/classification/convolution_based/_minirocket.py
- **RDST** (Guillaume, Vrain, and Elloumi 2022): Integrates dilation into random shapelet transform and includes the core mechanism of random convolutional kernels in Rocket.
  Source code: https://github.com/aeon-toolkit/aeon/blob/main/aeon/classification/shapelet_based/_rdst.py
- **MultiRocket-Hydra** (Dempster, Schmidt, and Webb 2023): Combines convolutional-based (MultiRocekt (Tan et al. 2022)) and dictionary-based methods through a pool of competing kernels for enriched feature diversity.
  Source code: https://github.com/aeon-toolkit/aeon/blob/main/aeon/classification/convolution_based/_mr_hydra.py
- **InceptionTime** (Ismail Fawaz et al. 2020): An ensemble of deep CNNs inspired by Inception-v4 architecture with varying filter lengths for multiscale feature extraction.
  Source code: https://github.com/timeseriesAI/tsai/blob/main/tsai/models/InceptionTime.py
- **TS2Vec** (Yue et al. 2022): Learns contextual timestamp-level representations via hierarchical contrastive learning across augmented subsequences.
  Source code: https://github.com/zhihanyue/ts2vec
- **PatchTST** (Nie et al. 2023): Adopts patch-based segmentation and channel-independent processing to enable efficient Transformer-based modeling of time series.
  Source code: https://github.com/yuqinie98/patchtst
- **TimesNet** (Wu et al. 2023): Transforms 1D time series into 2D representations using multi-period blocks and applies TimesBlock for generic temporal modeling.
  Source code: https://github.com/thuml/TimesNet
- **SoftShape** (Liu et al. 2025): Replaces hard shapelet selection with soft sparsification shapes and learning from inter-shape and intra-shape patterns for classification.
  Source code: https://github.com/qianlima-lab/SoftShape
- **GPT4TS** (Zhou et al. 2023): Leverages pre-trained large language models with patch-based and channel-independent designs for time series tasks.
  Source code: https://github.com/DAMO-DI-ML/One_Fits_All
- **MOMENT** (Goswami et al. 2024): Introduces high-capacity Transformer models pre-trained via masked prediction over a large heterogeneous time series dataset.
  Source code: https://github.com/MOMENT-timeseries-foundation-model/MOMENT
- **UniTS** (Gao et al. 2024): A unified multitask time series model based on Transformer that utilizes task tokenization to integrate predictive and generative tasks into a single framework.
  Source code: https://github.com/mims-harvard/UniTS
- **NuTime** (Lin et al. 2024): Incorporates numerically multi-scaled embeddings into a Transformer for pre-training to embed data of arbitrary numerical amplitude.
  Source code: https://github.com/chenguolin/NuTime
- **Mantis** (Feofanov et al. 2025): A foundation model for time series built upon the Vision Transformer and trained with a contrastive learning approach.
  Source code: https://github.com/vfeofanov/mantis
- **Chronos** (Ansari et al. 2024): A probabilistic foundation model for time series forecasting that tokenizes scaled and quantized series values into discrete vocabularies, enabling the use of transformer-based language architectures. Chronos is pretrained on large-scale real and synthetic datasets using cross-entropy loss to enhance generalization across domains.
  Source code: https://github.com/amazon-science/chronos-forecasting
- **Moirai** (Woo et al. 2024): A Transformer-based foundation model for time series forecasting that employs a masked encoder architecture and is pretrained on the large-scale open time series archive containing over 27

Table 1: The summary of the 128 UCR univariate time series datasets. "Train Size" and "Test Size" denote the number of samples in the respective splits. "No. Class" indicates the total number of target classes in each dataset, while "Length" refers to the sequence length of the univariate time series. Entries marked "Vary" in the length column indicate datasets containing samples with missing values.

ID	Dataset Name	Train Size	Test Size	No. Classes	Length	ID	Dataset Name	Train Size	Test Size	No. Classes	Length
1	ACSF1	100	100	10	1460	65	ItalyPowerDemand	67	1029	2	24
2	Adiac	390	391	37	176	66	LargeKitchenAppliances	375	375	3	720
3	AllGestureWiimoteX	300	700	10	Vary	67	Lightning2	60	61	2	637
4	AllGestureWiimoteY	300	700	10	Vary	68	Lightning7	70	73	7	319
5	AllGestureWiimoteZ	300	700	10	Vary	69	Mallat	55	2345	8	1024
6	ArrowHead	36	175	3	251	70	Meat	60	60	3	448
7	Beef	30	30	5	470	71	MedicalImages	381	760	10	99
8	BeetleFly	20	20	2	512	72	MelbournePedestrian	1194	2439	10	24
9	BirdChicken	20	20	2	512	73	MiddlePhalanxOutlineAgeGroup	400	154	3	80
10	BME	30	150	3	128	74	MiddlePhalanxOutlineCorrect	600	291	2	80
11	Car	60	60	4	577	75	MiddlePhalanxTW	399	154	6	80
12	CBF	30	900	3	128	76	MixedShapesRegularTrain	500	2425	5	1024
13	Chinatown	20	343	2	24	77	MixedShapesSmallTrain	100	2425	5	1024
14	ChlorineConcentration	467	3840	3	166	78	MoteStrain	20	1252	2	84
15	CinCECGTorso	40	1380	4	1639	79	NonInvasiveFetalECGThorax1	1800	1965	42	750
16	Coffee	28	28	2	286	80	NonInvasiveFetalECGThorax2	1800	1965	42	750
17	Computers	250	250	2	720	81	OliveOil	30	30	4	570
18	CricketX	390	390	12	300	82	OSULeaf	200	242	6	427
19	CricketY	390	390	12	300	83	PhalangesOutlinesCorrect	1800	858	2	80
20	CricketZ	390	390	12	300	84	Phoneme	214	1896	39	1024
21	Crop	7200	16800	24	46	85	PickupGestureWiimoteZ	50	50	10	Vary
22	DiatomSizeReduction	16	306	4	345	86	PigAirwayPressure	104	208	52	2000
23	DistalPhalanxOutlineAgeGroup	400	139	3	80	87	PigArtPressure	104	208	52	2000
24	DistalPhalanxOutlineCorrect	600	276	2	80	88	PigCVP	104	208	52	2000
25	DistalPhalanxTW	400	139	6	80	89	PLAID	537	537	11	Vary
26	DodgerLoopDay	78	80	7	288	90	Plane	105	105	7	144
27	DodgerLoopGame	20	138	2	288	91	PowerCons	180	180	2	144
28	DodgerLoopWeekend	20	138	2	288	92	ProximalPhalanxOutlineAgeGroup	400	205	3	80
29	Earthquakes	322	139	2	512	93	ProximalPhalanxOutlineCorrect	600	291	2	80
30	ECG200	100	100	2	96	94	ProximalPhalanxTW	400	205	6	80
31	ECG5000	500	4500	5	140	95	RefrigerationDevices	375	375	3	720
32	ECGFiveDays	23	861	2	136	96	Rock	20	50	4	2844
33	ElectricDevices	8926	7711	7	96	97	ScreenType	375	375	3	720
34	EOGHorizontalSignal	362	362	12	1250	98	SemgHandGenderCh2	300	600	2	1500
35	EOGVerticalSignal	362	362	12	1250	99	SemgHandMovementCh2	450	450	6	1500
36	EthanolLevel	504	500	4	1751	100	SemgHandSubjectCh2	450	450	5	1500
37	FaceAll	560	1690	14	131	101	ShakeGestureWiimoteZ	50	50	10	Vary
38	FaceFour	24	88	4	350	102	ShapeletSim	20	180	2	500
39	FacesUCR	200	2050	14	131	103	ShapesAll	600	600	60	512
40	FiftyWords	450	455	50	270	104	SmallKitchenAppliances	375	375	3	720
41	Fish	175	175	7	463	105	SmoothSubspace	150	150	3	15
42	FordA	3601	1320	2	500	106	SonyAIBORobotSurface1	20	601	2	70
43	FordB	3636	810	2	500	107	SonyAIBORobotSurface2	27	953	2	65
44	FreezerRegularTrain	150	2850	2	301	108	StarLightCurves	1000	8236	3	1024
45	FreezerSmallTrain	28	2850	2	301	109	Strawberry	613	370	2	235
46	Fungi	18	186	18	201	110	SwedishLeaf	500	625	15	128
47	GestureMidAirD1	208	130	26	Vary	111	Symbols	25	995	6	398
48	GestureMidAirD2	208	130	26	Vary	112	SyntheticControl	300	300	6	60
49	GestureMidAirD3	208	130	26	Vary	113	ToeSegmentation1	40	228	2	277
50	GesturePebbleZ1	132	172	6	Vary	114	ToeSegmentation2	36	130	2	343
51	GesturePebbleZ2	146	158	6	Vary	115	Trace	100	100	4	275
52	GunPoint	50	150	2	150	116	TwoLeadECG	23	1139	2	82
53	GunPointAgeSpan	135	316	2	150	117	TwoPatterns	1000	4000	4	128
54	GunPointMaleVersusFemale	135	316	2	150	118	UMD	36	144	3	150
55	GunPointOldVersusYoung	136	315	2	150	119	UWaveGestureLibraryAll	896	3582	8	945
56	Ham	109	105	2	431	120	UWaveGestureLibraryX	896	3582	8	315
57	HandOutlines	1000	370	2	2709	121	UWaveGestureLibraryY	896	3582	8	315
58	Haptics	155	308	5	1092	122	UWaveGestureLibraryZ	896	3582	8	315
59	Herring	64	64	2	512	123	Wafer	1000	6164	2	152
60	HouseTwenty	40	119	2	2000	124	Wine	57	54	2	234
61	InlineSkate	100	550	7	1882	125	WordSynonyms	267	638	25	270
62	InsectEPGRegularTrain	62	249	3	601	126	Worms	181	77	5	900
63	InsectEPGSmallTrain	17	249	3	601	127	WormsTwoClass	181	77	2	900
64	InsectWingbeatSound	220	1980	11	256	128	Yoga	300	3000	2	426

billion observations from nine domains.
Source code: https://github.com/SalesforceAIResearch/uni2ts

For both traditional and deep learning-based supervised classification methods, as well as foundation models, we employ the same 128 UCR time series datasets with identical preprocessing procedures and fixed train-test splits to ensure fair comparison. For deep learning-based time series foundation models, our objective is to evaluate their performance in fully supervised and zero-shot classification tasks, rather than comparing different pretraining strategies or source datasets. Therefore, using a consistent downstream dataset for evaluation is fair and reasonable. Specifically, UniShape, NuTime, and Mantis are all pretrained on the same 1.89M-sample dataset, confirming that their performance improvements primarily depend on architectural design. GPT4TS is based on the LLM-based GPT2 backbone, requiring approximately eight H100 GPUs for 24 hours of

Table 2: The summary of the 30 additional univariate time series datasets used in our zero-shot feature extraction experiments. Suffixes denote preprocessing strategies: _eq indicates unequal-length series standardized via mean-padding with Gaussian noise; _nmv refers to datasets originally containing missing values but filtered to retain only complete cases; _disc represents datasets sourced from the TSER archive (Tan et al. 2021). "Train Size" and "Test Size" denote the number of samples in the respective splits. "No. Class" indicates the total number of target classes in each dataset, while "Length" refers to the sequence length of the univariate time series. "Category" refers to the application domain of each dataset.

ID	Dataset Name	Train Size	Test Size	Length	No. Classes	Category
1	AconityMINIPrinterLarge_eq	2403	1184	300	2	Sensor
2	AconityMINIPrinterSmall_eq	589	292	300	2	Sensor
3	AllGestureWiimoteX_eq	300	700	500	10	Motion
4	AllGestureWiimoteY_eq	300	700	500	10	Motion
5	AllGestureWiimoteZ_eq	300	700	500	10	Motion
6	AsphaltObstaclesUni_eq	390	391	736	4	Sensor
7	AsphaltPavementTypeUni_eq	1055	1056	2371	3	Sensor
8	AsphaltRegularityUni_eq	751	751	4201	2	Sensor
9	Colposcopy	99	101	180	6	Image
10	Covid3Month_disc	140	140	84	3	Other
11	DodgerLoopDay_nmv	67	77	288	7	Sensor
12	DodgerLoopGame_nmv	17	127	288	2	Sensor
13	DodgerLoopWeekend_nmv	18	126	288	2	Sensor
14	ElectricDeviceDetection	624	3768	256	2	Image
15	FloodModeling1_disc	471	471	266	2	Simulated
16	FloodModeling2_disc	466	466	266	2	Simulated
17	FloodModeling3_disc	429	429	266	2	Simulated
18	GestureMidAirD1_eq	208	130	360	26	Motion
19	GestureMidAirD2_eq	208	130	360	26	Motion
20	GestureMidAirD3_eq	208	130	360	26	Motion
21	GesturePebbleZ1_eq	132	172	455	6	Motion
22	GesturePebbleZ2_eq	146	158	455	6	Motion
23	KeplerLightCurves	920	399	4767	7	Sensor
24	MelbournePedestrian_nmv	1138	2319	24	10	Sensor
25	PhoneHeartbeatSound	424	182	3053	5	Other
26	PickupGestureWiimoteZ_eq	50	50	361	10	Motion
27	PLAID_eq	537	537	1345	11	Device
28	ShakeGestureWiimoteZ_eq	50	50	385	10	Motion
29	SharePriceIncrease	965	966	60	2	Other
30	Tools	310	134	2926	5	Other

pretraining on text-based source datasets, while MOMENT and UniTS are pretrained on data aggregated from multiple heterogeneous time series sources. During fine-tuning, GPT4TS, MOMENT, and UniTS adopt the same settings as UniShape to assess the benefits of classification-specific pretraining data and strategies.

All baselines are implemented based on the above source code with author-provided recommended hyperparameters. Experiments are conducted in a unified Python environment on identical hardware to ensure fair and reproducible comparisons.

A.3 Implementation Details

For all baseline methods and *UniShape*, we report the average classification accuracy on the default test sets, computed over five independent runs with different random seeds. Notably, the 128 UCR time series datasets do not provide a dedicated validation set, which easily leads to overfitting for deep learning models. To mitigate this issue, *InceptionTime* (Ismail Fawaz et al. 2020) adopts an ensemble of five independently initialized models, demonstrating improved robustness on UCR time series datasets. Similarly, *NuTime* (Lin et al. 2024) shows that ensembling five identical pre-trained foundation models leads to further performance gains. In line with these insights, we employ a five-model ensemble for *UniShape* during fine-tuning across all UCR datasets to ensure reliable evaluation and minimize overfitting risks. For the baselines, InceptionTime, PatchTST, SoftShape, GPT4TS, MOMENT, UniTS, NuTime, and Mantis (excluding TimesNet due to its extensive runtime requirements) employ a five-model ensemble on the 128 UCR datasets for downstream classification evaluation to obtain the final test results. Across all 128 UCR datasets, we follow the standard practice of using the original training/test splits for both *UniShape* and all baseline methods. For deep learning-based approaches, following the setting described

in (Wang, Yan, and Oates 2017; Early et al. 2024), we report test performance using the model checkpoint that achieves the lowest training loss. Algorithm 1 presents the detailed pseudo-code implementation of the UniShape model.

During the pre-training phase, *UniShape* employs equal window and stride lengths ($W_q = K_q$) across five temporal resolutions: $\{64, 32, 16, 8, 4\}$. However, our preliminary experimental results show that directly fine-tuning across all five scales is both computationally intensive and time-consuming. Moreover, selecting an optimal scale via a held-out validation set yields notable improvements in downstream classification performance. Additionally, fixing the stride to $K_q = 4$ (or $K_q = 2$ for sequences shorter than 64) during fine-tuning enhances the density of extracted subsequences and contributes to improved accuracy. To balance computational efficiency with performance, we adopt the validation strategy proposed in prior work (Grabocka et al. 2014; Eldele et al. 2024), in which 20% of the original training set is randomly sampled as a validation set. This validation subset is used to identify the optimal scale W_q for each UCR dataset with minimal training epoch, while maintaining a fixed stride of $K_q = 4$ or 2 throughout fine-tuning.

For fine-tuning on the 128 UCR time series datasets, unlike NuTime and Mantis which standardize all sequences to a length of 512 via interpolation, UniShape uses the original sequence lengths. This approach better preserves the interpretability of learned shapelets in classification results. Our preliminary experiments show that fine-tuning UniShape with original sequence lengths yields comparable overall performance to using interpolated 512-length sequences across the 128 UCR datasets, while reducing runtime for sequences shorter than 512. Furthermore, larger batch sizes, such as 512, consistently improve classification performance during fine-tuning, whereas smaller batch sizes, such as 8 or 16, may be preferable for datasets with fewer than 100 samples. Accordingly, for UniShape, batch sizes are selected using a validation split from the training set for each UCR time series dataset, with 512 recommended as the default settings.

Algorithm 1: The pseudo-code of UniShape Model.

Require: The shape-aware adapter f_{adapter}, Transformer encoder f_{encoder}, classification head f_{cls}, pre-train epochs E_{pre}, and fine-tune epochs E_{fine}.
Ensure: Trained f_{adapter}, f_{encoder}, and f_{cls}.

Stage 1: Pre-training on the Source Dataset
for $i = 1$ to E_{pre} **do**
  Sample multiscale subsequences as inputs (Eq. (1))
  Extract tokens $\mathbf{c}^{(Q)}$, $\mathbf{z}^{(Q)}$ via f_{adapter} (Eq. (2)–(4))
  Encode tokens: $(\mathbf{c}^{(Q)\prime}, \mathbf{z}^{(Q)\prime}) \leftarrow f_{\text{encoder}}(\mathbf{c}^{(Q)}, \mathbf{z}^{(Q)})$
  Compute pre-train loss $\mathcal{L}_{\text{pretrain}}$ (Eq. (7)–(11))
  Update f_{adapter}, f_{encoder} via backpropagation
end for
Obtain weights frozen f_{adapter} and f_{encoder}
Stage 2: Fine-tuning on Target Datasets
for each target dataset $\mathcal{D}_t$ **do**
  Using frozen f_{adapter} and f_{encoder} as initialization
  for $j = 1$ to E_{fine} **do**
    Sample subsequences from $\mathcal{D}_t$ using Eq. (1)
    Extract $\mathbf{c}^{(Q)}$, $\mathbf{z}^{(Q)}$ via f_{adapter}
    Forward to encoder f_{encoder}: $(\mathbf{c}^{(Q)\prime}, \mathbf{z}^{(Q)\prime})$
    Compute cross-entropy loss $\mathcal{L}_{\text{ce}}$ (Eq. (12)) via f_{cls}
    Compute total loss $\mathcal{L}_{\text{finetune}}$ (Eq. (8),(12),(13))
    Update f_{adapter}, f_{cls} via backpropagation
  end for
  Evaluate final model on test set of $\mathcal{D}_t$
end for
return f_{adapter}, f_{encoder}, f_{cls}

B Details of Overall Evaluation Results

Table 3 presents the test classification accuracy of UniShape and 14 baseline methods across all 128 UCR time series datasets under the fully supervised setting. To enhance clarity and readability, we report only the test classification accuracy for each method on each dataset, omitting standard deviations across multiple runs. The same presentation strategy is followed in subsequent experimental sections.

C Details of Pretraining Analysis Results

This section examines the impact of different pre-training configurations on the downstream classification performance of UniShape. Specifically, we investigate the effects of varying the size and label ratio of the pre-training dataset. Table 4 reports the test classification accuracy of UniShape after pre-training on four progressively larger datasets: UCR, UEA, UCR+UEA, and the full dataset (ALL) containing 1.89 million time series samples. The downstream evaluation is conducted on 18 selected UCR datasets under the fully supervised fine-tuning setting.

Table 5 evaluates the effect of label ratio of samples in the pre-training phase. Here, we use the entire UCR archive as the pre-training source and vary the label availability ratio from 0% to 100% (i.e., 0%, 1%, 10%, 50%, and 100%).

D Details of Zero-Shot Results

In this section, we evaluate the feature extraction capability of various time series foundation models in a zero-shot setting. We select task-agnostic models (GPT4TS, MOMENT, UniTS), forecasting-based models (Chronos, Moirai), and classification-based models (NuTime, Mantis) as baselines. Since NuTime and Mantis standardize input sequences to length 512 via interpolation, and some datasets have original sequence lengths shorter than 64, making it difficult for UniShape to utilize all sliding window sizes, we interpolate all sequences from the 30 additional univariate time series datasets to length 512 for consistent model input. For UniShape, pretraining with temporal scales $\{64, 32, 16, 8\}$ yields better zero-shot classification performance than including scale 4. Accordingly, in this evaluation, UniShape retains $\{64, 32, 16, 8\}$ for pretraining and uses the four window sizes in one model for zero-shot feature extraction, resulting in approximately 2.8 million parameters. All stride sizes are set to $K_q = 4$, using a single model for all datasets. Features extracted from the 30 datasets are used to

train a Random Forest classifier, enabling assessment of the foundation models' representation quality without any task-specific fine-tuning. Table 6 presents the zero-shot classification accuracies achieved by UniShape and baseline models, thereby highlighting their generalization ability across unseen datasets.

E Details of Ablation Study Results

This section presents a comprehensive ablation study analyzing the effect of different training paradigms and architectural components on UniShape's downstream classification performance. Given the high computational cost of pre-training on the full 1.89 million sample dataset (approximately 13 hours using four NVIDIA RTX A6000 GPUs) and the impracticality of performing exhaustive ablation across all 128 UCR datasets, we adopt a more efficient setup. Based on the findings in Table 4, increasing the scale of the pre-training data improves downstream performance, and using the UCR archive alone is sufficient for meaningful representation learning.

Following Table 4, we therefore restrict our ablation to the UCR archive for pre-training and perform fine-tuning on a representative subset of 18 UCR datasets. Table 7 reports the test classification accuracy of UniShape under different ablation configurations, revealing the contribution of each design choice to the overall model performance.

F Hyperparameter and Runtime Analysis

For the hyperparameter analysis, all experiments are conducted using the same pre-training dataset and the same set of 18 UCR datasets used in Table 5, ensuring consistency across evaluations and reducing computational cost.

Effect of Shape Token Selection Ratio ϵ. Table 8 presents the test accuracy of UniShape when varying the shape token ratio ϵ, which determines the proportion of tokens with the highest attention scores used in the shape-prototype contrastive loss during fine-tuning. As shown in Table 8, UniShape achieves the lowest average rank across the 18 UCR datasets when $\epsilon = 0.6$. Given that analyzing different token ratios during pre-training is computationally expensive, we set $\epsilon = 0.6$ as the default value throughout all UniShape experiments in pre-training and fine-tuning stages.

Effect of Prototype-based Pre-training Loss Weight λ. Table 9 shows the impact of the weight coefficient λ in Equation (9) of the main text, which controls the contribution of the shape-prototype and instance-prototype contrastive loss during pre-training. The results indicate that setting $\lambda = 0.01$ leads to the best average rank performance across the 18 datasets. Therefore, we fix $\lambda = 0.01$ in all subsequent UniShape pre-training stages.

Effect of Shape-Prototype Loss Weight μ for Fine-tuning. Table 10 shows the classification performance of UniShape under different values of μ, the loss weight associated with the shape-prototype contrastive loss defined in Equation (13) of the main text. When $\mu = 0.01$, UniShape achieves the best average rank, suggesting that this value provides an effective balance when applying shape-prototype loss as an auxiliary objective during fine-tuning. Consequently, we set $\mu = 0.01$ for all fine-tuning experiments involving this auxiliary loss.

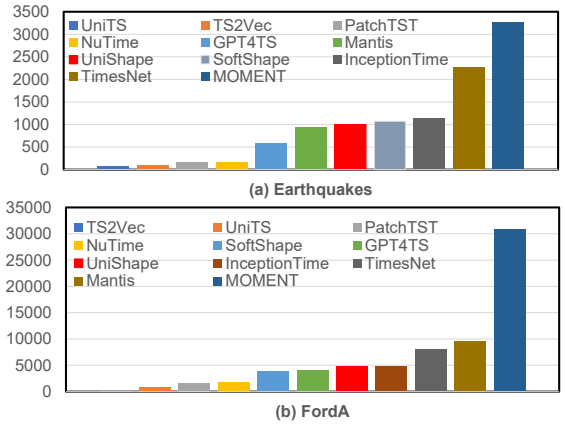

Figure 1: Runtime analysis (training and evaluation) of deep learning-based baselines on the *Earthquakes* and *FordA* datasets from the UCR archive.

Other Hyperparameters. For the MOMENTum coefficient β used in Equation (6) to update the class prototypes, we adopt the default value $\beta = 0.9$.

Runtime Analysis. In the pretraining stage, UniShape requires approximately 13 hours on four NVIDIA RTX A6000 GPUs using 1.89 million samples. For fine-tuning on target domains, we compare its runtime with ten representative deep learning-based TSC models. Methods such as Rocket, MiniRocket, RDST, and MR-H were excluded, as they use randomly initialized convolutional kernels without gradient-based optimization, resulting in minimal training time. The evaluation was carried out on two datasets with differing training sizes: *Earthquakes* (322 samples) and *FordA* (3,601 samples). As shown in Figure 1, UniShape requires less runtime (seconds) than InceptionTime, TimesNet, and MOMENT across both datasets, and is also more efficient than Mantis on the larger *FordA* dataset. Notably, UniShape employs the same ensemble strategy as InceptionTime, fine-tuning five independently pre-trained models. The reported runtime reflects the total cost of this ensemble; in practice, using a single model could reduce training time by approximately a factor of five. While UniShape is marginally slower than baselines such as GPT4TS and NuTime, it offers substantially improved interpretability by learning shapelets.

ID	Dataset Name	Rocket	MiniRocket	RDST	MR-H	InceptionTime	TS2Vec	PatchTST	TimesNet	SoftShape	GPT4TS	MOMENT	UniTS	NuTime	Mantis	UniShape
1	ACSF1	0.8600	0.9100	0.9300	0.8600	**0.9400**	0.8800	0.7300	0.3000	0.8800	0.6100	0.5000	0.7700	0.6700	0.7000	0.8000
2	Adiac	0.7903	0.8235	0.7136	**0.8344**	0.8107	0.7545	0.6675	0.6292	0.8159	0.7826	0.1458	0.5115	0.8031	0.8159	0.8289
3	AllGestureWiimoteX	0.7629	0.7429	0.7471	0.7600	0.7229	0.6729	0.2643	0.4529	0.7871	0.5029	0.5157	0.6329	0.7086	**0.7914**	0.7743
4	AllGestureWiimoteY	0.7486	0.7571	0.7386	0.7886	0.7943	0.6771	0.2529	0.4557	0.7643	0.5129	0.5986	0.6443	0.7557	**0.8243**	0.7929
5	AllGestureWiimoteZ	**0.7886**	0.7314	0.7100	0.7529	0.7771	0.7371	0.1757	0.4071	0.7443	0.4071	0.4886	0.5700	0.6614	0.7586	0.7429
6	ArrowHead	0.8057	0.8400	0.8686	0.8629	0.8400	0.8400	0.6971	0.7486	0.8457	0.8057	0.5657	0.7543	0.8343	0.8686	**0.8857**
7	Beef	0.8000	0.8667	0.8333	0.8667	0.7333	0.7000	0.7333	0.6000	**0.9200**	0.8000	0.5333	0.7333	0.9167	0.7333	0.8667
8	BeetleFly	0.9000	0.8500	**0.9500**	0.8400	0.9000	0.8500	0.7000	0.6500	0.9000	0.8500	**0.9500**	0.7000	0.7500	0.9000	0.8500
9	BirdChicken	0.9000	0.9000	0.9000	0.9000	0.9000	0.8000	0.9000	0.8000	**1.0000**	0.5500	0.8500	0.8000	0.9000	**1.0000**	0.9500
10	BME	**1.0000**	0.9933	0.9933	**1.0000**	**1.0000**	0.9800	0.9733	0.9933	0.9867	0.9533	0.8800	0.9733	0.9933	**1.0000**	**1.0000**
11	Car	0.9000	0.9167	**0.9333**	0.9233	0.8833	0.7833	0.7167	0.7667	0.8667	0.8333	0.6667	0.7833	0.8333	0.8833	**0.9333**
12	CBF	**1.0000**	0.9989	0.9922	0.9956	0.9956	**1.0000**	0.7489	0.9356	0.9978	0.8278	0.9700	0.8767	0.9878	**1.0000**	**1.0000**
13	Chinatown	0.9826	0.9855	0.9855	0.9497	0.9855	0.9768	0.9072	**0.9884**	0.9826	0.9652	0.9855	0.9768	0.9797	0.9855	0.9565
14	ChlorineConcentration	0.8156	0.7630	0.7576	0.8193	0.8654	0.8281	0.6523	0.8229	**0.8971**	0.8242	0.5664	0.8578	0.7630	0.7633	0.8539
15	CinCECGTorso	0.8304	0.8688	0.9335	0.9288	0.6638	0.7993	0.7457	0.8159	0.9036	0.7558	0.5565	0.8058	**0.9609**	0.8449	0.9500
16	Coffee	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	0.9643	**1.0000**	**1.0000**	**1.0000**	0.8929	**1.0000**	**1.0000**	**1.0000**	**1.0000**
17	Computers	0.7720	0.7360	0.7360	0.7618	0.7880	0.6600	0.5080	0.5680	0.6560	0.5760	0.6480	0.5480	**0.7920**	0.7400	0.7680
18	CricketX	**0.8282**	0.8103	0.8026	0.8026	0.7923	0.7923	0.5205	0.5385	0.8256	0.5333	0.6949	0.6872	0.7821	0.8179	**0.8282**
19	CricketY	**0.8564**	0.8385	0.8308	0.8251	0.7974	0.7564	0.6026	0.5538	0.8051	0.5615	0.6564	0.6692	0.8026	0.8308	0.8333
20	CricketZ	**0.8590**	0.8282	0.8462	0.8385	0.8282	0.8128	0.5615	0.6026	0.8462	0.5564	0.7308	0.6949	0.7974	**0.8590**	0.8436
21	Crop	0.7574	0.7622	0.7637	0.7678	0.7727	0.7540	0.0766	0.7197	0.7733	0.6665	0.6605	0.6548	**0.7888**	0.7877	0.7725
22	DiatomSizeReduction	0.9771	0.9281	0.9248	0.9641	0.9150	0.9869	**0.9902**	0.9150	0.9771	0.9542	0.7222	0.9118	0.9085	0.9739	0.9706
23	DistalPhalanxOutlineAgeGroup	0.7410	0.7338	**0.7842**	0.7626	0.7482	0.7122	0.7338	0.6763	0.7770	0.7482	0.6691	0.7122	0.7266	0.7050	0.7698
24	DistalPhalanxOutlineCorrect	0.7609	0.7681	0.7717	0.7616	0.7536	0.7572	0.6848	0.7681	0.7790	0.7391	0.6775	0.7246	0.7609	0.7681	**0.7899**
25	DistalPhalanxTW	0.7050	0.6763	**0.7266**	0.6978	0.7122	0.6978	0.6331	0.6547	0.6835	0.6043	0.6619	0.6547	0.6835	0.6547	0.7194
26	DodgerLoopDay	0.6125	**0.6875**	0.6750	0.5750	0.6000	0.5500	0.4250	0.2500	0.6500	0.4500	0.3000	0.5125	0.5125	0.5750	0.6250
27	DodgerLoopGame	0.8696	0.8768	0.8478	0.8986	**0.9348**	0.8986	0.6522	0.5145	0.8696	0.6884	0.8333	0.8261	0.8623	0.8406	0.9203
28	DodgerLoopWeekend	0.9710	**0.9855**	0.9783	0.9755	0.9783	0.9638	0.9348	0.7609	0.9783	0.9710	0.9493	0.9710	0.9710	0.9783	**0.9855**
29	Earthquakes	0.7554	**0.7626**	0.7338	0.7338	0.7554	0.7482	0.7266	0.7554	0.7482	0.7410	0.7482	0.6978	0.7482	0.7482	0.7554
30	ECG200	0.9051	0.9100	0.9000	0.9100	0.9100	0.8700	0.8400	0.8300	0.9150	0.8500	0.8700	0.8800	0.8000	0.9000	**0.9200**
31	ECG5000	0.9478	0.9449	0.9429	0.9460	0.9453	0.9404	0.9209	0.8993	0.9436	0.9309	0.9424	0.9376	0.9440	0.9422	**0.9498**
32	ECGFiveDays	**1.0000**	**1.0000**	0.9977	**1.0000**	**1.0000**	**1.0000**	0.7689	0.8920	0.9965	0.9547	0.8281	0.8211	0.9768	0.9570	**1.0000**
33	ElectricDevices	0.7312	0.7496	0.7408	0.7424	0.7256	0.7255	0.3167	0.6457	0.6676	0.5774	0.6388	0.6090	**0.7506**	0.7373	0.7248
34	EOGHorizontalSignal	0.6492	0.5856	0.6768	0.6354	0.6326	0.5552	0.4586	0.3094	0.6298	0.2680	0.4558	0.4724	0.6657	0.6768	**0.6796**
35	EOGVerticalSignal	0.5470	0.5470	0.5497	0.5138	0.4834	0.4420	0.3398	0.2624	**0.5608**	0.2541	0.3978	0.5083	0.5552	0.5359	0.5193
36	EthanolLevel	0.5800	0.6000	0.6600	0.6380	**0.8520**	0.4620	0.4260	0.5140	0.8100	0.6740	0.2640	0.5600	0.6340	0.4460	0.8500
37	FaceAll	**0.9485**	0.8077	0.7941	0.8077	0.8041	0.7669	0.4876	0.6799	0.8006	0.7479	0.6982	0.7467	0.7976	0.8178	0.8994
38	FaceFour	0.9773	**0.9886**	0.9773	0.9545	0.8864	0.8977	0.5341	0.8182	0.9659	0.7614	0.8636	0.6818	0.9545	**0.9886**	0.9545
39	FacesUCR	0.9615	0.9595	**0.9629**	0.9595	0.9473	0.9341	0.6902	0.8317	0.9624	0.8307	0.7956	0.8098	0.9210	0.9522	0.9620
40	FiftyWords	0.8242	0.8418	0.8484	0.8415	0.7473	0.7648	0.6440	0.5824	0.8308	0.6879	0.7714	0.7143	0.7890	**0.8637**	0.8462
41	Fish	0.9829	0.9771	**0.9943**	0.9829	0.9771	0.9429	0.7543	0.7600	0.9714	0.8800	0.7543	0.8800	0.9600	0.9771	0.9886
42	FordA	0.9311	0.9470	0.9470	0.9553	0.9515	0.9348	0.7841	0.9136	0.9606	0.8818	0.9174	0.9311	0.9348	0.9439	**0.9629**
43	FordB	0.8049	0.8198	0.8160	0.8309	0.8580	0.7938	0.5716	0.7519	**0.8679**	0.7395	0.7802	0.7716	0.8074	0.8160	0.8605
44	FreezerRegularTrain	0.9979	**0.9996**	0.9989	**0.9996**	0.9905	0.9814	0.9825	0.9628	0.9909	0.9523	0.7747	0.9926	0.9944	0.9958	0.9972
45	FreezerSmallTrain	0.9523	0.9716	0.9944	**0.9951**	0.8832	0.8986	0.9004	0.7337	0.8154	0.6744	0.7733	0.6375	0.9530	0.8140	0.9828
46	Fungi	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	0.9355	0.3978	0.8656	**1.0000**	0.7312	0.9785	0.8011	0.8065	0.9946	**1.0000**
47	GestureMidAirD1	0.7462	0.7308	0.7615	0.7846	0.7538	0.6615	0.3462	0.4538	0.7538	0.5615	0.6692	0.6846	0.7462	**0.8077**	0.8000
48	GestureMidAirD2	0.6615	0.6385	0.6538	0.7000	0.6154	0.4846	0.1462	0.3615	0.6308	0.4615	0.5462	0.4769	0.6846	0.6923	**0.7231**
49	GestureMidAirD3	0.4615	0.4769	0.5308	**0.5462**	0.3615	0.3769	0.1846	0.2923	0.3692	0.3615	0.2846	0.3769	**0.5462**	0.5154	0.4692
50	GesturePebbleZ1	0.9012	0.8895	0.9012	0.9128	0.8837	0.5581	0.5349	0.7558	0.9186	0.6686	0.7616	0.8663	0.9012	0.9244	**0.9477**
51	GesturePebbleZ2	0.8924	0.9051	0.8987	0.8671	0.9051	0.5127	0.5063	0.5886	0.9494	0.5759	0.8038	0.8924	0.8354	0.8861	**0.9747**
52	GunPoint	**1.0000**	0.9933	**1.0000**	**1.0000**	**1.0000**	0.9800	0.9400	0.9667	**1.0000**	0.9600	0.9267	0.9667	0.9933	**1.0000**	**1.0000**
53	GunPointAgeSpan	0.9968	0.9968	0.9968	**1.0000**	0.9747	0.9557	0.7342	0.8797	0.9810	0.9494	0.9177	0.9209	0.9873	0.9968	0.9968
54	GunPointMaleVersusFemale	**1.0000**	**1.0000**	**1.0000**	**1.0000**	0.9968	0.9937	0.9905	0.9905	0.9842	0.9968	0.9778	0.9937	**1.0000**	**1.0000**	**1.0000**
55	GunPointOldVersusYoung	0.9873	**1.0000**	**1.0000**	**1.0000**	0.7587	**1.0000**	**1.0000**	0.9365	0.9746	0.9175	0.8667	0.9460	**1.0000**	**1.0000**	**1.0000**
56	Ham	0.7429	0.6952	0.7429	0.7429	0.7429	0.7524	0.6000	0.6952	**0.7714**	0.6762	0.6095	0.7333	0.6952	0.7333	0.7619
57	HandOutlines	0.9432	0.9432	0.9514	0.9459	0.9514	0.9189	0.9135	0.8649	0.9324	0.8892	0.8108	0.9054	0.9270	0.9486	**0.9649**
58	Haptics	0.5455	0.5195	**0.5747**	0.5260	0.4935	0.5130	0.4091	0.4383	0.4805	0.4383	0.4091	0.4481	0.5195	0.4968	0.5682
59	Herring	0.6250	0.6875	0.6875	0.7344	0.6250	0.5625	0.6250	0.4844	0.5938	0.5625	0.6406	0.6094	0.5312	0.5469	**0.7969**
60	HouseTwenty	0.9580	0.9664	0.9664	**0.9916**	0.9832	0.7899	0.6891	0.6639	0.9580	0.6050	0.9244	0.8235	0.9076	0.9832	0.9832
61	InlineSkate	0.4764	0.4764	0.4436	**0.4964**	0.4327	0.3836	0.3236	0.2309	0.3764	0.2982	0.2255	0.2927	0.4055	0.4164	0.4909
62	InsectEPGRegularTrain	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	0.8996	0.7470	**1.0000**	**1.0000**	**1.0000**
63	InsectEPGSmallTrain	0.9839	0.9960	0.9960	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	0.9357	**1.0000**	0.8755	0.7269	**1.0000**	**1.0000**	**1.0000**
64	InsectWingbeatSound	0.6697	**0.6702**	0.6566	0.6572	0.6298	0.6338	0.5328	0.6232	0.6646	0.6167	0.6247	0.6071	0.6480	0.6470	0.6682
65	ItalyPowerDemand	**0.9689**	0.9640	0.8756	0.9679	0.9670	0.9572	0.9252	0.9514	0.9640	0.9553	0.9349	0.9572	0.9572	0.9543	0.9640
66	LargeKitchenAppliances	0.9040	0.8720	0.8453	0.9088	0.9067	0.8667	0.5493	0.5227	0.8773	0.5040	0.7600	0.7280	0.8640	0.8640	**0.9253**
67	Lightning2	0.7869	0.7541	0.7541	0.7213	**0.8689**	0.8525	0.7049	0.6393	0.7869	0.6721	0.7213	0.7705	0.7541	0.8197	0.8525
68	Lightning7	0.8082	0.7808	0.7945	0.8219	0.8082	0.8493	0.4932	0.7123	0.8493	0.7534	0.7260	0.6575	0.7397	0.8082	**0.8630**
69	Mallat	0.9561	0.9450	0.9258	0.9288	0.9629	0.8772	0.8333	0.9011	0.9557	0.8789	0.4307	0.8985	0.9441	0.9608	**0.9659**
70	Meat	0.9500	0.9500	**0.9667**	0.9333	0.9167	**0.9667**	0.6500	0.9000	0.9333	0.9000	0.6667	0.8833	0.9333	0.7167	0.9500
71	MedicalImages	0.8000	0.8013	0.7658	0.7974	0.8000	0.7895	0.7039	0.6908	0.7276	0.6987	0.7211	0.6868	0.7868	0.7961	**0.8039**
72	MelbournePedestrian	0.9122	0.9616	0.9645	0.9690	0.9216	0.9514	0.4498	0.9224	0.9000	0.9155	0.7976	0.8710	**0.9784**	0.9731	0.9702
73	MiddlePhalanxOutlineAgeGroup	0.5649	0.5714	0.5974	0.5649	0.5519	0.6429	0.5519	0.5844	0.5519	0.4870	**0.6558**	0.4610	0.4870	0.5065	0.5714
74	MiddlePhalanxOutlineCorrect	0.8282	0.8419	0.8247	0.8488	0.8419	0.8076	0.7354	0.7869	0.8351	0.7698	0.5739	0.7904	0.8213	0.8213	**0.8660**
75	MiddlePhalanxTW	0.5455	0.5455	0.5325	0.5130	0.5325	**0.6039**	0.5519	0.5455	0.5519	0.5649	0.5974	0.4805	0.4740	0.5065	0.5974
76	MixedShapesRegularTrain	0.9720	0.9707	0.9810	**0.9823**	0.9452	0.9241	0.6177	0.7885	0.9472	0.8610	0.7744	0.8854	0.9724	0.9773	0.9753
77	MixedShapesSmallTrain	0.9365	0.9509	**0.9687**	0.9507	0.8874	0.8623	0.7728	0.5744	0.8858	0.7955	0.7023	0.8041	0.9460	0.9579	0.9509
78	MoteStrain	0.9161	0.9281	0.9217	0.9457	0.9089	0.8570	0.8690	0.8922	0.9081	0.8786	0.8986	0.8770	**0.9561**	0.9377	0.9305
79	NonInvasiveFetalECGThorax1	0.9542	0.9496	0.9252	**0.9644**	0.9191	0.9288	0.8952	0.8229	0.9511	0.8926	0.6153	0.9079	0.9354	0.9267	0.9511
80	NonInvasiveFetalECGThorax2	**0.9684**	0.9598	0.9394	**0.9684**	0.9201	0.9425	0.9232	0.8555	0.9466	0.9084	0.6911	0.9064	0.9491	0.9344	0.9517
81	OliveOil	0.9333	0.9333	0.9333	**0.9667**	0.7000	0.9000	0.4000	0.7667	0.8667	0.7667	0.4000	0.5000	0.7333	0.8667	0.8667
82	OSULeaf	0.9421	0.9504	0.9815	0.9628	0.9174	0.8512	0.5165	0.5041	0.8678	0.5165	0.7438	0.5909	0.9380	0.9545	**0.9835**
83	PhalangesOutlinesCorrect	0.8275	0.8333	0.8089	0.8450	0.8508	0.8263	0.7075	0.7879	0.8298	0.7727	0.6340	0.7389	0.8159	0.8263	**0.8590**
84	Phoneme	0.2811	0.2911	0.3370	0.3539	0.3402	0.3085	0.1218	0.0860	0.3133	0.1160	0.3017	0.1313	0.3238	0.3576	**0.3729**
85	PickupGestureWiimoteZ	0.8000	**0.8800**	0.8400	0.8200	0.8000	0.6200	0.6200	0.5800	0.8600	0.5400	0.6600	0.7200	0.7400	0.8000	0.8600
86	PigAirwayPressure	0.4038	**0.8510**	0.8269	0.7115	0.1202	0.3269	0.0433	0.0337	0.1154	0.0673	0.0577	0.0769	0.5913	0.4279	0.3510
87	PigArtPressure	0.9567	**0.9856**	0.9519	0.9567	0.8317	0.1490	0.1154	0.0192	0.9663	0.0769	0.1154	0.1346	0.8846	0.9038	**0.9856**
88	PigCVP	0.9423	**0.9519**	0.9327	0.8942	0.8837	0.6827	0.0577	0.0288	0.7933	0.1010	0.0865	0.0529	0.8750	0.8558	0.7500
89	PLAID	0.5791	**0.9441**	0.9199	0.9311	0.7831	0.5289	0.2197	0.5289	0.5698	0.4916	0.3575	0.5009	0.9050	0.8864	0.8939

ID	Dataset Name	Rocket	MiniRocket	RDST	MR-H	InceptionTime	TS2Vec	PatchTST	TimesNet	SoftShape	GPT4TS	MOMENT	UniTS	NuTime	Mantis	UniShape
90	Plane	**1.0000**	**1.0000**	**1.0000**	**1.0000**	0.9905	0.9905	0.9619	0.9714	**1.0000**	0.9810	0.9810	0.9714	**1.0000**	**1.0000**	0.9905
91	PowerCons	0.9333	0.9944	0.9944	0.9833	0.9889	0.9556	**1.0000**	0.9944	0.9778	0.9944	0.8722	0.9889	**1.0000**	**1.0000**	**1.0000**
92	ProximalPhalanxOutlineAgeGroup	0.8585	0.8439	0.8488	0.8634	0.8585	0.8341	0.8098	0.8634	0.8537	0.8537	0.8488	0.8488	**0.8732**	0.8293	0.8585
93	ProximalPhalanxOutlineCorrect	0.8969	0.9038	0.8935	0.9278	0.9107	0.8625	0.7732	0.9175	0.8900	0.8660	0.7560	0.8557	0.8797	0.8454	**0.9347**
94	ProximalPhalanxTW	0.7951	0.7951	0.8293	0.8049	0.7854	0.8244	0.7707	0.8000	0.7951	0.8000	0.7561	0.7659	0.7659	0.7902	**0.8341**
95	RefrigerationDevices	0.5280	0.4773	0.5680	0.5173	0.5253	0.6000	0.3493	0.4160	0.5360	0.4187	0.5173	0.4160	0.5653	0.5227	**0.6667**
96	Rock	**0.9000**	0.8200	0.8800	0.8600	0.6200	0.6800	0.6800	0.7600	0.8600	0.8600	0.6400	0.8400	**0.9000**	0.8600	0.8600
97	ScreenType	0.4613	0.4853	0.5360	0.5600	**0.5840**	0.4053	0.3440	0.4160	0.4720	0.3920	0.4347	0.3813	0.5680	0.4667	0.5653
98	SemgHandGenderCh2	0.9350	0.8933	0.9267	0.9500	0.9200	0.8600	0.7883	0.8600	0.9250	0.8283	0.7483	0.8883	0.9250	0.9350	**0.9600**
99	SemgHandMovementCh2	0.6289	0.6978	0.6956	0.7733	0.7022	0.6156	0.3000	0.3022	0.6422	0.3222	0.4222	0.4733	0.8400	**0.9067**	0.8422
100	SemgHandSubjectCh2	0.8889	0.8689	0.8511	0.9022	0.8956	0.7444	0.2689	0.3822	0.8800	0.5489	0.6444	0.8556	0.9067	**0.9422**	0.9333
101	ShakeGestureWiimoteZ	0.9200	0.9400	0.9000	0.9400	0.9400	0.8800	0.8400	0.7400	0.9600	0.5400	0.8800	0.8600	0.9000	**0.9800**	0.9400
102	ShapeletSim	**1.0000**	**1.0000**	**1.0000**	**1.0000**	0.9167	0.9833	0.6000	0.4833	**1.0000**	0.4833	0.9667	0.5444	**1.0000**	**1.0000**	**1.0000**
103	ShapesAll	0.9050	0.9200	0.9233	**0.9283**	0.8117	0.9033	0.3133	0.5783	0.8850	0.7050	0.6717	0.7633	0.9083	0.9050	0.9050
104	SmallKitchenAppliances	0.8240	0.8293	0.8213	0.8187	0.7760	0.7253	0.3973	0.5787	0.7733	0.5360	0.6800	0.5760	0.7840	0.7893	**0.8427**
105	SmoothSubspace	0.9733	0.9600	**0.9933**	0.9800	0.9867	0.9400	0.9467	0.9533	**0.9933**	0.9000	0.7200	0.8467	0.9800	**0.9933**	0.9800
106	SonyAIBORobotSurface1	0.9235	0.8968	0.9068	0.8869	0.8286	0.8802	0.7937	0.6789	0.7953	0.7072	0.7754	0.6705	0.8120	0.7987	**0.9368**
107	SonyAIBORobotSurface2	0.9234	0.9224	0.8940	**0.9323**	0.8846	0.8709	0.8069	0.8237	0.8395	0.8636	0.9014	0.8279	0.8877	0.8930	0.8909
108	StarLightCurves	0.9808	**0.9826**	**0.9826**	0.9814	0.9780	0.9701	0.9216	0.9303	0.9749	0.9618	0.8568	0.9596	0.9781	0.9771	0.9815
109	Strawberry	0.9811	**0.9838**	0.9784	0.9757	0.9757	0.9595	0.8541	0.9703	**0.9838**	0.9595	0.8189	0.9703	0.9730	0.9649	0.9811
110	SwedishLeaf	0.9664	0.9664	0.9632	0.9760	0.9536	0.9280	0.9376	0.8512	0.9504	0.8912	0.8576	0.8992	0.9664	0.9728	**0.9776**
111	Symbols	0.9739	0.9849	0.9809	0.9809	0.9809	0.9678	0.8101	0.8452	0.9688	0.8040	0.9578	0.8784	0.9849	0.9889	**0.9940**
112	SyntheticControl	**1.0000**	0.9800	0.9967	0.9967	**1.0000**	0.9967	0.7900	0.9667	**1.0000**	0.9533	0.9967	0.9667	0.9867	0.9967	**1.0000**
113	ToeSegmentation1	0.9561	0.9561	0.9518	0.9561	**0.9605**	0.9430	0.7237	0.6711	0.9430	0.5789	0.9254	0.7719	0.9386	0.9474	0.9561
114	ToeSegmentation2	0.9308	0.9154	0.9077	0.9154	0.9231	0.9308	0.5154	0.7077	0.9385	0.6462	0.9000	0.7692	0.9077	0.9385	**0.9538**
115	Trace	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	0.8600	**1.0000**	0.8400	0.9600	**1.0000**	**1.0000**	**1.0000**	**1.0000**
116	TwoLeadECG	0.9991	0.9982	0.9965	0.9982	0.9982	0.9807	0.8595	0.7392	0.9535	0.8648	0.7480	0.8824	0.9842	0.9965	**1.0000**
117	TwoPatterns	**1.0000**	0.9973	**1.0000**	**1.0000**	**1.0000**	0.9998	0.9995	0.9618	**1.0000**	0.9542	0.9908	0.9990	**1.0000**	**1.0000**	**1.0000**
118	UMD	0.9931	0.9931	**1.0000**	0.9931	0.9931	**1.0000**	0.9722	**1.0000**	0.9861	0.9653	0.8958	0.9653	0.9931	0.9931	**1.0000**
119	UWaveGestureLibraryAll	0.9765	0.9735	**0.9802**	0.9799	0.9037	0.9372	0.9199	0.9238	0.9710	0.9196	0.8939	0.9534	0.9704	0.9788	0.9796
120	UWaveGestureLibraryX	0.8565	0.8479	0.8554	0.8713	0.8102	0.8035	0.7457	0.6965	0.8378	0.7563	0.8138	0.7892	0.8691	**0.8864**	0.8786
121	UWaveGestureLibraryY	0.7786	0.7781	0.7641	0.8085	0.7610	0.7264	0.6563	0.6119	0.7560	0.6309	0.6924	0.6991	0.8051	0.8185	**0.8269**
122	UWaveGestureLibraryZ	0.7956	0.8009	0.7998	0.8222	0.7803	0.7688	0.6924	0.6189	0.7892	0.6454	0.7367	0.7125	0.8130	0.8236	**0.8361**
123	Wafer	0.9981	0.9994	**1.0000**	0.9997	0.9982	0.9966	0.9943	0.9955	0.9959	0.9951	0.9927	0.9942	0.9989	0.9989	0.9997
124	Wine	0.8148	0.8333	0.8704	0.9059	0.6111	0.8519	0.5000	0.7963	0.6481	**0.9259**	0.5000	0.5741	0.7222	0.6296	0.7963
125	WordSynonyms	0.7508	0.7524	0.7492	0.7774	0.6536	0.6818	0.5470	0.5188	0.7429	0.5549	0.5690	0.6066	0.7038	**0.7837**	0.7492
126	Worms	0.7273	0.7532	0.6883	0.7792	0.7532	0.7273	0.4935	0.3247	0.7792	0.4545	0.5714	0.5584	0.7143	0.7532	**0.7922**
127	WormsTwoClass	0.8182	0.7792	0.7662	0.7792	0.7922	0.8312	0.5195	0.4416	0.8052	0.5974	0.7792	0.6234	0.7273	0.8052	**0.8831**
128	Yoga	0.9177	0.9050	0.9277	**0.9290**	0.8997	0.8833	0.7963	0.7843	0.8860	0.7927	0.6573	0.7833	0.8497	0.8907	0.9197
	Avg. Acc	0.8487	0.8545	0.8571	0.8621	0.8315	0.8016	0.6500	0.6897	0.8388	0.7100	0.7020	0.7357	0.8353	0.8441	**0.8708**
	Avg. Rank	4.87	4.84	4.80	3.97	6.10	8.32	12.69	11.83	5.89	11.69	12.10	11.42	6.68	5.21	**2.71**
	P-value	7.80E-06	3.25E-03	8.58E-03	2.96E-02	2.55E-11	3.12E-11	2.12E-26	3.00E-24	3.68E-32	7.62E-23	7.04E-25	3.20E-23	2.08E-10	1.69E-06	-

Table 3: The detailed test classification accuracy results comparison on 128 UCR datasets under the fully supervised setting. The best results are in **bold**.

Dataset Name	UCR	UEA	UCR+UEA	ALL
ArrowHead	0.8344	**0.8857**	0.8515	**0.8857**
CBF	**1.0000**	**1.0000**	**1.0000**	**1.0000**
CricketX	0.8279	0.7974	**0.8381**	0.8282
DistalPhalanxOutlineAgeGroup	0.7494	0.7482	0.7322	**0.7698**
DistalPhalanxOutlineCorrect	0.7781	0.7717	0.7709	**0.7899**
ECG5000	0.9438	0.9416	0.9478	**0.9498**
EOGVerticalSignal	0.5695	0.5746	**0.5912**	0.5193
EthanolLevel	0.8500	0.8050	**0.8620**	0.8520
Fish	**1.0000**	0.9771	0.9885	0.9886
GunPoint	0.9933	0.9867	**1.0000**	**1.0000**
InsectWingbeatSound	0.6606	0.6545	0.6667	**0.6682**
ItalyPowerDemand	0.9546	0.9534	0.9459	**0.9640**
MelbournePedestrian	0.9447	0.9555	0.9596	**0.9702**
MiddlePhalanxTW	0.5525	0.5779	0.5850	**0.5974**
MixedShapesRegularTrain	0.9793	0.9819	**0.9835**	0.9753
OSULeaf	0.9629	0.9917	**0.9959**	0.9835
Trace	**1.0000**	**1.0000**	**1.0000**	**1.0000**
WordSynonyms	0.7508	**0.7618**	0.7602	0.7492
Avg. Acc	0.8529	0.8536	0.8599	**0.8606**
Win	3	4	8	**11**

Table 4: The detailed classification accuracy results comparison on 18 UCR datasets with different numbers of pre-training samples. *UEA* contains about 1.39 million samples, *UCR+UEA* contains about 1.45 million samples, and *ALL* contains about 1.89 million samples. *Win* denotes number of datasets where the method achieves the best accuracy. The best results are in **bold**.

Dataset Name	r = 0%	r = 1%	r = 10%	r = 50%	r = 100%
ArrowHead	0.7715	0.7872	0.8344	0.8358	**0.8915**
CBF	0.9978	**1.0000**	**1.0000**	0.9978	**1.0000**
CricketX	0.8279	0.8304	0.8279	0.8409	**0.8438**
DistalPhalanxOutlineAgeGroup	0.7266	0.7422	**0.7494**	0.7250	0.7278
DistalPhalanxOutlineCorrect	0.7717	0.7781	0.7781	0.7817	**0.7854**
ECG5000	0.9418	0.9460	0.9438	**0.9501**	0.9491
EOGVerticalSignal	0.5944	0.5392	0.5695	0.5972	**0.6275**
EthanolLevel	0.8340	0.8300	0.8500	**0.8640**	0.8600
Fish	**1.0000**	0.9772	**1.0000**	**1.0000**	0.9829
GunPoint	0.9866	0.9866	**0.9933**	**0.9933**	0.9866
InsectWingbeatSound	0.6278	0.6404	0.6606	0.6722	**0.6868**
ItalyPowerDemand	0.9536	**0.9663**	0.9546	0.9478	0.9634
MelbournePedestrian	0.9386	**0.9590**	0.9447	0.9578	0.9239
MiddlePhalanxTW	0.5200	0.5265	**0.5525**	**0.5525**	0.4875
MixedShapesRegularTrain	0.9684	0.9715	0.9793	**0.9855**	0.9845
OSULeaf	0.9381	0.9340	0.9629	0.9463	**0.9670**
Trace	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
WordSynonyms	0.7125	0.7225	0.7508	0.7852	**0.7899**
Avg. Acc	0.8395	0.8410	0.8529	0.8574	**0.8588**
P-value	1.87E-02	3.28E-02	2.00E-01	4.10E-01	-

Table 5: The detailed test classification accuracy results comparison on 18 UCR datasets with different training sample labeling ratios for pre-training. And r denotes the labeling ratio of pre-training samples. The best results are in **bold**.

ID	Dataset Name	RandomForest	GPT4TS	MOMENT	Chronos	Moirai	UniTS	NuTime	Mantis	UniShape
1	AconityMINIPrinterLarge_eq	0.9096	0.6596	0.8660	**0.9265**	0.8049	0.5819	0.8438	0.8710	0.8691
2	AconityMINIPrinterSmall_eq	0.9486	0.7158	**0.9589**	0.9315	0.8562	0.6575	0.8562	0.9315	0.9486
3	AllGestureWiimoteX_eq	0.5286	0.4186	0.5343	0.4100	0.2929	0.0800	**0.5414**	0.4943	0.5343
4	AllGestureWiimoteY_eq	0.5571	0.4157	**0.5857**	0.3829	0.3014	0.0971	0.5129	0.5086	0.5771
5	AllGestureWiimoteZ_eq	0.4643	0.3871	0.5286	0.4343	0.2614	0.1029	0.4543	0.4757	**0.5557**
6	AsphaltObstaclesUni_eq	0.7366	0.7417	0.7647	0.8159	0.4731	0.2379	0.7954	0.7980	**0.8235**
7	AsphaltPavementTypeUni_eq	0.7794	0.7992	0.7670	0.6960	0.4830	0.3598	**0.8239**	0.7367	0.7860
8	AsphaltRegularityUni_eq	0.9228	0.9414	0.9174	0.8935	0.6591	0.4807	0.9534	0.9534	**0.9627**
9	Colposcopy	0.4059	0.4059	0.3762	**0.4257**	0.3267	0.3168	0.4158	**0.4257**	0.3960
10	Covid3Month_disc	0.6393	0.6066	0.5410	**0.6885**	**0.6885**	0.4754	0.6066	0.6721	0.5738
11	DodgerLoopDay_nmv	**0.6883**	0.1818	0.3377	0.4286	0.4026	0.1948	0.4286	0.4286	0.5455
12	DodgerLoopGame_nmv	0.7638	0.6063	0.7717	0.7165	**0.8346**	0.4803	0.6772	0.7044	0.6614
13	DodgerLoopWeekend_nmv	**0.9841**	0.4841	0.9444	0.7540	0.7460	0.5476	0.9524	0.9562	0.9206
14	ElectricDeviceDetection	0.8758	0.8559	0.8559	0.8646	**0.8766**	0.8299	0.8697	0.8612	0.8607
15	FloodModeling1_disc	0.8119	0.7475	0.7673	0.7871	0.7624	0.7030	**0.8218**	0.7475	0.7871
16	FloodModeling2_disc	**0.9353**	0.9303	0.9254	0.9303	0.9254	0.9204	0.9303	0.9254	0.9254
17	FloodModeling3_disc	**0.7609**	0.6957	0.6793	0.6630	0.6685	0.6304	0.7065	0.6576	0.6793
18	GestureMidAirD1_eq	0.5154	0.3000	0.6231	0.4231	0.2846	0.0385	0.4846	0.5077	**0.6615**
19	GestureMidAirD2_eq	0.4154	0.0923	0.6015	0.5769	0.2154	0.1000	0.4769	0.4923	**0.6077**
20	GestureMidAirD3_eq	0.2231	0.1615	**0.4308**	0.2462	0.1923	0.0385	0.2538	0.2846	0.3923
21	GesturePebbleZ1_eq	0.6570	0.6453	0.7500	0.6337	0.4767	0.1628	0.7442	0.8447	**0.8488**
22	GesturePebbleZ2_eq	0.6646	0.5190	0.7785	0.6962	0.5823	0.1329	0.7848	**0.8608**	0.7722
23	KeplerLightCurves	0.7143	0.2982	**0.9349**	0.8897	0.8296	0.2682	0.7268	0.7945	0.8296
24	MelbournePedestrian_nmv	0.8906	0.7861	0.6270	0.8185	0.7628	0.1035	0.8952	0.8323	**0.8991**
25	PhoneHeartbeatSound	0.5989	0.5824	0.6429	**0.6978**	0.6209	0.5165	0.6538	0.6923	**0.6978**
26	PickupGestureWiimoteZ_eq	0.6800	0.5200	0.5200	0.6800	0.3400	0.0600	0.6800	0.7800	**0.8000**
27	PLAID_eq	**0.7933**	0.4562	0.7263	0.7858	0.5605	0.1415	0.6648	0.6909	0.7393
28	ShakeGestureWiimoteZ_eq	0.7800	0.7200	0.8400	0.8000	0.5600	0.1200	0.8800	**0.8900**	0.8600
29	SharePriceIncrease	0.6884	**0.6915**	0.6781	0.6801	0.6718	0.6387	0.6894	0.6822	0.6822
30	Tools	0.4552	0.4328	0.6418	**0.7015**	0.6119	0.2761	0.6269	0.6567	0.5896
	Avg. Acc	0.6930	0.5600	0.6972	0.6793	0.5691	0.3431	0.6917	0.7052	**0.7262**
	Avg. Rank	3.77	6.37	4.17	4.10	6.37	8.90	3.53	3.67	**3.07**
	P-value	2.57E-02	1.79E-06	3.98E-02	4.91E-03	5.69E-06	6.24E-10	3.36E-03	3.15E-02	-

Table 6: The detailed test classification accuracy results comparison on 30 additional time series datasets with a zero-shot feature extraction. The best results are in **bold**.

Dataset Name	UniShape	w/o pre-training			Pre-training and Fine-tuning							
		Encoder			Adapter			Prototype			Encoder	
		using Trans	re CNN	re MLP	w/o ada	re Trans	re MLP	w/o Ins	w/o Shape	w/o All	re CNN	re MLP
ArrowHead	**0.8344**	0.7144	0.8001	0.8058	0.8058	0.7955	0.7829	0.7944	0.8172	0.7658	0.8115	0.7144
CBF	**1.0000**	0.9978	0.9844	0.9967	**1.0000**	0.9967	0.9956	**1.0000**	**1.0000**	**1.0000**	0.9956	0.9778
CricketX	0.8279	0.8484	0.7920	0.6048	0.8016	0.8305	0.8228	0.8433	0.8305	0.8433	**0.8538**	0.3253
DistalPhalanxOutlineAgeGroup	**0.7494**	0.7206	0.7422	0.6847	0.7422	0.7134	0.7366	**0.7494**	0.7350	0.7350	0.7062	0.6127
DistalPhalanxOutlineCorrect	0.7781	0.7854	0.8035	0.7709	0.7999	0.7754	**0.8052**	0.7817	0.7607	0.7590	0.7926	0.7455
ECG5000	0.9438	0.9311	0.9381	0.9418	0.9423	0.9412	0.9409	0.9385	**0.9441**	0.9434	0.9403	0.9321
EOGVerticalSignal	0.5695	0.5553	0.5249	0.2929	0.5629	0.5332	0.5529	0.5498	0.5553	0.5346	**0.5774**	0.0829
EthanolLevel	0.8500	0.8640	0.8320	0.3540	0.8560	0.8700	**0.8800**	0.8330	0.8320	0.8240	0.8240	0.3500
Fish	**1.0000**	0.9486	0.9828	0.6571	0.9943	0.9863	0.9828	0.9943	0.9886	0.9860	0.9960	0.1486
GunPoint	0.9933	0.9867	**1.0000**	0.9933	0.9233	0.9833	0.9721	0.9933	0.9933	0.9933	0.9933	0.9933
InsectWingbeatSound	**0.6606**	0.6444	0.6485	0.4646	0.6470	0.6521	0.6348	0.6510	0.6502	0.6444	0.6602	0.4939
ItalyPowerDemand	0.9546	**0.9692**	0.9595	0.9566	0.9466	0.9604	0.9517	0.9468	0.9366	0.9359	0.9585	0.9585
MelbournePedestrian	0.9447	0.8727	0.9486	0.0931	0.9368	0.9470	0.9388	0.9251	0.9445	0.9417	**0.9560**	0.1964
MiddlePhalanxTW	0.5525	0.5520	0.5395	**0.5914**	0.5420	0.5460	0.5070	0.5330	0.5330	0.5395	0.5405	0.5785
MixedShapesRegularTrain	**0.9793**	0.9732	0.9587	0.9253	0.9694	0.9715	0.9707	0.9682	0.9777	0.9740	0.9759	0.2408
OSULeaf	0.9629	0.9505	0.7480	0.8100	0.9549	0.9587	0.9629	0.9670	**0.9753**	0.9629	0.9735	0.5125
Trace	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
WordSynonyms	0.7508	0.7430	0.6725	0.3543	0.7446	0.7414	0.7383	0.7305	**0.7712**	0.7571	0.7669	0.3088
Avg. Acc	**0.8529**	0.8365	0.8264	0.6832	0.8428	0.8446	0.8431	0.8444	0.8470	0.8411	0.8512	0.5651
Avg. Rank	**3.00**	6.33	6.61	8.11	5.50	5.72	6.78	5.56	4.39	5.89	3.94	9.61
P-value	-	2.96E-02	2.33E-02	3.35E-03	1.65E-02	1.68E-02	2.70E-02	5.92E-03	2.11E-02	8.15E-03	3.40E-01	4.11E-04

Table 7: The detailed classification accuracy results of the ablation study on 18 UCR time series datasets. *w/o* means without, and *re* means replace. The best results are in **bold**.

Dataset Name	ϵ=0	ϵ=0.2	ϵ=0.4	ϵ=0.6	ϵ=0.8	ϵ=1
ArrowHead	0.8016	0.7901	0.8244	**0.8344**	0.8187	0.8016
CBF	**1.0000**	0.9989	0.9989	**1.0000**	0.9989	0.9989
CricketX	0.8253	0.8228	0.8228	**0.8279**	0.8228	0.8202
DistalPhalanxOutlineAgeGroup	0.7538	0.7554	0.7538	0.7494	**0.7682**	0.7466
DistalPhalanxOutlineCorrect	0.7238	0.7681	0.7745	**0.7781**	**0.7781**	0.7528
ECG5000	0.9437	0.9301	0.9417	0.9438	0.9424	**0.9439**
EOGVerticalSignal	0.5667	0.5816	0.5540	0.5695	**0.5833**	0.5750
EthanolLevel	0.8620	**0.8900**	0.8640	0.8500	0.8200	0.8640
Fish	0.9930	0.9930	0.9987	**1.0000**	0.9872	0.9987
GunPoint	**0.9933**	**0.9933**	**0.9933**	**0.9933**	**0.9933**	**0.9933**
InsectWingbeatSound	0.6394	0.6434	0.6500	**0.6606**	0.6555	0.6565
ItalyPowerDemand	0.9469	0.9585	0.9556	0.9546	**0.9624**	0.9556
MelbournePedestrian	**0.9876**	0.9321	0.9647	0.9447	0.9300	0.9786
MiddlePhalanxTW	0.5510	0.5575	0.5510	0.5525	**0.5640**	0.5510
MixedShapesRegularTrain	0.9789	**0.9793**	**0.9793**	**0.9793**	**0.9793**	**0.9793**
OSULeaf	0.9629	**0.9670**	**0.9670**	0.9629	0.9629	0.9629
Trace	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
WordSynonyms	**0.7540**	0.7346	0.7377	0.7508	0.7508	0.7461
Avg. Acc	0.8491	0.8498	0.8517	**0.8529**	0.8510	0.8514
Avg. Rank	3.39	3.06	2.89	**2.33**	2.61	2.94

Table 8: The detailed test classification accuracy results on 18 UCR time series datasets with different top ratio ϵ of shape tokens for fine-tuning. The best results are in **bold**.

Dataset Name	$\lambda = 0.001$	$\lambda = 0.01$	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.5$	$\lambda = 0.9$	$\lambda = 0.99$	$\lambda = 0.999$
ArrowHead	0.8515	0.8344	0.8630	**0.8744**	0.8344	0.8344	0.8344	0.8344
CBF	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	0.9978	**1.0000**	0.9989
CricketX	0.8177	0.8279	0.8459	0.8177	0.8125	0.8254	**0.8561**	0.8202
DistalPhalanxOutlineAgeGroup	**0.8429**	0.7494	0.7206	0.7853	0.8069	0.8141	0.7853	0.7925
DistalPhalanxOutlineCorrect	0.7709	0.7781	0.7600	**0.7926**	0.7527	0.7672	0.7383	0.7238
ECG5000	**0.9476**	0.9438	0.9443	0.9420	0.9383	0.9447	0.9434	0.9458
EOGVerticalSignal	0.5115	**0.5695**	0.5501	0.5557	0.5363	0.5560	0.5667	0.5612
EthanolLevel	0.8600	0.8500	0.8220	0.8680	0.8240	0.8620	0.8400	**0.8800**
Fish	0.9943	**1.0000**	0.9828	0.9886	0.9714	0.9886	0.9886	0.9886
GunPoint	0.9800	**0.9933**	**0.9933**	0.9866	**0.9933**	**0.9933**	**0.9933**	0.9866
InsectWingbeatSound	0.6591	**0.6606**	0.6358	0.6298	0.6368	0.6535	0.6353	0.6439
ItalyPowerDemand	0.9206	0.9546	0.9410	0.9313	0.9332	0.9566	**0.9614**	0.9498
MelbournePedestrian	0.9402	**0.9447**	0.9235	0.8459	0.8827	0.9406	0.9198	0.8382
MiddlePhalanxTW	0.5480	0.5525	0.5330	0.5590	0.5135	0.5655	**0.5785**	0.5720
MixedShapesRegularTrain	0.9905	0.9793	0.9913	0.9880	0.9859	**0.9917**	0.9851	0.9847
OSULeaf	**0.9711**	0.9629	0.9133	0.9629	0.9546	0.9670	0.9546	0.9670
Trace	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
WordSynonyms	0.7195	**0.7508**	0.6881	0.6725	0.7054	0.6803	0.7070	0.7226
Avg. Acc	0.8514	**0.8529**	0.8393	0.8445	0.8379	0.8522	0.8493	0.8450
Avg. Rank	3.56	**3.00**	4.61	4.44	5.28	3.17	3.72	4.06

Table 9: The detailed test classification accuracy results on 18 UCR time series datasets with different weighting hyperparameter λ of prototype-based contrastive loss for pre-training. The best results are in **bold**.

Dataset Name	$\mu = 0.0001$	$\mu = 0.001$	$\mu = 0.01$	$\mu = 0.1$	$\mu = 0.5$	$\mu = 1$
ArrowHead	0.8317	0.8286	0.8344	**0.8515**	0.8344	0.8315
CBF	1.0000	**1.0011**	1.0000	**1.0011**	**1.0011**	**1.0011**
CricketX	0.8382	0.8330	0.8279	0.8305	0.8433	**0.8459**
DistalPhalanxOutlineAgeGroup	0.7062	0.6990	**0.7494**	**0.7494**	0.7206	**0.7494**
DistalPhalanxOutlineCorrect	**0.7998**	0.7672	0.7781	0.7781	0.7491	0.7962
ECG5000	0.9309	0.9413	**0.9438**	0.9409	0.9409	0.9258
EOGVerticalSignal	0.5695	0.5556	0.5695	0.5446	**0.5777**	0.5556
EthanolLevel	0.8460	0.8320	**0.8500**	0.8320	0.8440	0.8340
Fish	**1.0000**	0.9943	**1.0000**	**1.0000**	**1.0000**	0.9943
GunPoint	**0.9933**	**0.9933**	**0.9933**	**0.9933**	**0.9933**	**0.9933**
InsectWingbeatSound	0.6486	0.6571	**0.6606**	0.6516	0.6495	0.6390
ItalyPowerDemand	0.9400	0.9400	**0.9546**	0.9303	0.9420	0.9459
MelbournePedestrian	0.9365	0.9185	**0.9447**	0.9238	0.9108	0.9047
MiddlePhalanxTW	0.5655	0.5590	0.5525	0.5525	0.5525	**0.5850**
MixedShapesRegularTrain	0.9789	**0.9793**	**0.9793**	**0.9793**	0.9789	0.9776
OSULeaf	0.9629	0.9629	0.9629	0.9670	0.9629	**0.9876**
Trace	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
WordSynonyms	0.7571	0.7539	0.7508	0.7524	0.7539	**0.7618**
Avg. Acc	0.8503	0.8453	**0.8529**	0.8488	0.8475	0.8516
Avg. Rank	2.89	3.33	**2.28**	2.89	2.83	3.00

Table 10: The detailed test classification accuracy results on 18 UCR time series datasets with different weighting hyperparameter μ of shape-prototype contrastive loss for fine-tuning. The best results are in **bold**.

References

Andrzejak, R. G.; Lehnertz, K.; Mormann, F.; Rieke, C.; David, P.; and Elger, C. E. 2001. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E*, 64(6): 061907.

Anguita, D.; Ghio, A.; Oneto, L.; Parra, X.; Reyes-Ortiz, J. L.; et al. 2013. A public domain dataset for human activity recognition using smartphones. In *Esann*, volume 3, 3–4.

Ansari, A. F.; Stella, L.; Turkmen, A. C.; Zhang, X.; Mercado, P.; Shen, H.; Shchur, O.; Rangapuram, S. S.; Arango, S. P.; Kapoor, S.; et al. 2024. Chronos: Learning the Language of Time Series. *Transactions on Machine Learning Research*.

Bagnall, A.; Dau, H. A.; Lines, J.; Flynn, M.; Large, J.; Bostrom, A.; Southam, P.; and Keogh, E. 2018. The UEA multivariate time series classification archive, 2018. *arXiv preprint arXiv:1811.00075*.

Chen, X.; Xie, S.; and He, K. 2021. An empirical study of training self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, 9640–9649.

Clifford, G. D.; Liu, C.; Moody, B.; Lehman, L.-w. H.; Silva, I.; Li, Q.; Johnson, A. E.; and Mark, R. G. 2017. AF classification from a short single lead ECG recording: The PhysioNet/computing in cardiology challenge 2017. In *2017 computing in cardiology (CinC)*, 1–4. IEEE.

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.

Dempster, A.; Petitjean, F.; and Webb, G. I. 2020. ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels. *Data Mining and Knowledge Discovery*, 34(5): 1454–1495.

Dempster, A.; Schmidt, D. F.; and Webb, G. I. 2021. Minirocket: A very fast (almost) deterministic transform for time series classification. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, 248–257.

Dempster, A.; Schmidt, D. F.; and Webb, G. I. 2023. Hydra: Competing convolutional kernels for fast and accurate time series classification. *Data Mining and Knowledge Discovery*, 37(5): 1779–1805.

Early, J.; Cheung, G.; Cutajar, K.; Xie, H.; Kandola, J.; and Twomey, N. 2024. Inherently Interpretable Time Series Classification via Multiple Instance Learning. In *The Twelfth International Conference on Learning Representations*.

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. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, 2352–2359. International Joint Conferences on Artificial Intelligence Organization.

Eldele, E.; Ragab, M.; Chen, Z.; Wu, M.; and Li, X. 2024. TSLANet: rethinking transformers for time series representation learning. In *Proceedings of the 41st International Conference on Machine Learning*, 12409–12428.

Feofanov, V.; Wen, S.; Alonso, M.; Ilbert, R.; Guo, H.; Tiomoko, M.; Pan, L.; Zhang, J.; and Redko, I. 2025. Mantis: Lightweight calibrated foundation model for user-friendly time series classification. *arXiv preprint arXiv:2502.15637*.

Gao, S.; Koker, T.; Queen, O.; Hartvigsen, T.; Tsiligkaridis, T.; and Zitnik, M. 2024. Units: A unified multi-task time series model. *Advances in Neural Information Processing Systems*, 37: 140589–140631.

Goldberger, A. L.; Amaral, L. A.; Glass, L.; Hausdorff, J. M.; Ivanov, P. C.; Mark, R. G.; Mietus, J. E.; Moody, G. B.; Peng, C.-K.; and Stanley, H. E. 2000. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *circulation*, 101(23): e215–e220.

Goswami, M.; Szafer, K.; Choudhry, A.; Cai, Y.; Li, S.; and Dubrawski, A. 2024. MOMENT: A Family of Open Time-series Foundation Models. In *International Conference on Machine Learning*, 16115–16152. PMLR.

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.

Guillaume, A.; Vrain, C.; and Elloumi, W. 2022. Random dilated shapelet transform: A new approach for time series shapelets. In *International Conference on Pattern Recognition and Artificial Intelligence*, 653–664. Springer.

Ismail Fawaz, H.; Lucas, B.; Forestier, G.; Pelletier, C.; Schmidt, D. F.; Weber, J.; Webb, G. I.; Idoumghar, L.; Muller, P.-A.; and Petitjean, F. 2020. Inceptiontime: Finding alexnet for time series classification. *Data Mining and Knowledge Discovery*, 34(6): 1936–1962.

Kemp, B.; Zwinderman, A. H.; Tuk, B.; Kamphuisen, H. A.; and Oberye, J. J. 2000. Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG. *IEEE Transactions on Biomedical Engineering*, 47(9): 1185–1194.

Lessmeier, C.; Kimotho, J. K.; Zimmer, D.; and Sextro, W. 2016. Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification. In *PHM society European conference*, volume 3.

Lin, C.; Wen, X.; Cao, W.; Huang, C.; Bian, J.; Lin, S.; and Wu, Z. 2024. NuTime: Numerically Multi-Scaled Embedding for Large-Scale Time-Series Pretraining. *Transactions on Machine Learning Research*.

Liu, J.; Zhong, L.; Wickramasuriya, J.; and Vasudevan, V. 2009. uWave: Accelerometer-based personalized gesture recognition and its applications. *Pervasive and Mobile Computing*, 5(6): 657–675.

Liu, Z.; Luo, Y.; Li, B.; Eldele, E.; Wu, M.; and Ma, Q. 2025. Learning Soft Sparse Shapes for Efficient Time-Series Classification. In *Forty-second International Conference on Machine Learning*.

Ma, Q.; Liu, Z.; Zheng, Z.; Huang, Z.; Zhu, S.; Yu, Z.; and Kwok, J. T. 2024. A Survey on Time-Series Pre-Trained Models. *IEEE Transactions on Knowledge and Data Engineering*, 36(12): 7536–7555.

Middlehurst, M.; Schäfer, P.; and Bagnall, A. 2024. Bake off redux: a review and experimental evaluation of recent time series classification algorithms. *Data Mining and Knowledge Discovery*, 38(4): 1958–2031.

Nie, Y.; Nguyen, N. H.; Sinthong, P.; and Kalagnanam, J. 2023. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers. In *The Eleventh International Conference on Learning Representations*.

Tan, C. W.; Bergmeir, C.; Petitjean, F.; and Webb, G. I. 2021. Time series extrinsic regression: Predicting numeric values from time series data. *Data Mining and Knowledge Discovery*, 35(3): 1032–1060.

Tan, C. W.; Dempster, A.; Bergmeir, C.; and Webb, G. I. 2022. MultiRocket: multiple pooling operators and transformations for fast and effective time series classification. *Data Mining and Knowledge Discovery*, 36(5): 1623–1646.

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.

Woo, G.; Liu, C.; Kumar, A.; Xiong, C.; Savarese, S.; and Sahoo, D. 2024. Unified Training of Universal Time Series Forecasting Transformers. In *International Conference on Machine Learning*, 53140–53164. PMLR.

Wu, H.; Hu, T.; Liu, Y.; Zhou, H.; Wang, J.; and Long, M. 2023. TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis. In *The Eleventh International Conference on Learning Representations*.

Yue, Z.; Wang, Y.; Duan, J.; Yang, T.; Huang, C.; Tong, Y.; and Xu, B. 2022. Ts2vec: Towards universal representation of time series. In *Proceedings of the AAAI conference on artificial intelligence*, 8980–8987.

Zhang, X.; Zhao, Z.; Tsiligkaridis, T.; and Zitnik, M. 2022. Self-Supervised Contrastive Pre-Training For Time Series via Time-Frequency Consistency. In *Advances in neural information processing systems*, 1–16.

Zhou, T.; Niu, P.; Sun, L.; Jin, R.; et al. 2023. One fits all: Power general time series analysis by pretrained lm. *Advances in neural information processing systems*, 36: 43322–43355.