# TimeMIL: Advancing Multivariate Time Series Classification via a Time-aware Multiple Instance Learning

# Author

## Xiwen Chen

팔로우

Clemson University
g.clemson.edu의 이메일 확인됨 - 홈페이지
VLM/LLM  Post-training  Time Series Analysis  Computer Vision

| 제목 | 인용 | 연도 |
|---|---|---|
| Wildland fire detection and monitoring using a drone-collected rgb/ir image dataset<br>X Chen, B Hopkins, H Wang, L O'Neill, F Afghah, A Razi, P Fulé, J Coen, ...<br>IEEE Access 10, 121301-121317 | 180 | 2022 |
| Deep DIH: single-shot digital in-line holography reconstruction by deep learning<br>H Li, X Chen, Z Chi, C Mann, A Razi<br>Ieee Access 8, 202648-202659 | 79 | 2020 |
| Energy optimization for HVAC systems in multi-VAV open offices: A deep reinforcement learning approach<br>H Wang, X Chen, N Vital, E Duffy, A Razi<br>Applied Energy 356, 122354 | 67 | 2024 |
| Deep learning serves traffic safety analysis: A forward‐looking review<br>A Razi(1), X Chen(1), H Li, H Wang, B Russo, Y Chen, H Yu<br>IET Intelligent Transport Systems 17 (1), 22-71 | 64 | 2023 |
| SelfReg-UNet: Self-Regularized UNet for Medical Image Segmentation<br>W Zhu, X Chen, P Qiu, M Farazi, A Sotiras, A Razi, Y Wang<br>MICCAI2024 | 37 | 2024 |
| DH-GAN: a physics-driven untrained generative adversarial network for holographic imaging<br>X Chen, H Wang, A Razi, M Kozicki, C Mann<br>Optics Express 31 (6), 10114-10135 | 37 | 2023 |
| Visiongpt: Llm-assisted real-time anomaly detection for safe visual navigation<br>H Wang, J Qin, A Bastola, X Chen, J Suchanek, Z Gong, A Razi<br>arXiv preprint arXiv:2403.12415 | 33 | 2024 |
| TimeMIL: Advancing Multivariate Time Series Classification via a Time-aware Multiple Instance Learning<br>X Chen, P Qiu, W Zhu, H Li, H Wang, A Sotiras, Y Wang, A Razi<br>ICML2024 | 27 | 2024 |

SEOULTECH

# Introduction

- 시계열 데이터에서 관심 패턴은 일반적으로 국소적
- 표준 MIL은 인스턴스를 독립적인 개체로만 취급하므로, 인스턴스 간의 상호 관계는 고려하지 못하는 한계가 존재
- MTSC 데이터는 시간 순서가 중요하며, 순서를 무시하는 표준 MIL을 그대로 적용하기 어려움
- 다변량 및 다중 클래스 시계열 분류를 위한 MIL 프레임워크를 공식적으로 정형화

# Related Work

- 대부분 모델들은 supervised learning에 기반하고 있으며, 주로 아키텍처 디자인을 최적화하는 데 중점
- 최신 MIL 방법은 딥러닝을 이용해 자동으로 특징을 추출하고 종합하며, 전통적인 방식보다 성능이 좋음
- 최신 MIL 방법은 딥러닝을 이용해 자동으로 특징을 추출하고 종합하며, 전통적인 방식보다 성능이 좋음
  그 중 Attention-based MIL은 인스턴스의 중요도에 따라 가중치를 부여하여 해석 가능성을 제공
- ABMIL은 각 인스턴스를 독립적인 것으로 가정하기 때문에, 시간 순서에 따른 관계를 파악하는 데는 한계가 존재

# MTSC - Binary

- 하나라도 양성 인스턴스 레이블을 가지는 경우에만 양성으로 분류

$$y_i = \begin{cases} 0, \text{ iff } \sum_{t=1}^{T} y_i^t = 0, \ y_i^t \in \{0, 1\} \\ 1, \text{ otherwise,} \end{cases}$$

- MIL의 bag 예측()은 스코어 함수에 의해 확률로 계산

$$\boldsymbol{X}_i = \{\boldsymbol{x}_i^1, \cdots, \boldsymbol{x}_i^T\}$$

$$\hat{y}_i = S(\boldsymbol{X}_i)$$

# Multi-Class MTSC

- 일대다(one-vs-rest) 이진 MIL 수행
- 각 클래스별로 독립적인 Binary MIL을 수행

$$y_{i,c} = \begin{cases} 0, \text{ iff } \sum_{t=1}^{T} y_{i,c}^{t} = 0, \ y_{i,c}^{t} \in \{0,1\} \\ 1, \text{ otherwise,} \end{cases}$$

- 최종 bag 예측()은 가장 높은 확률을 가진 클래스로 계산

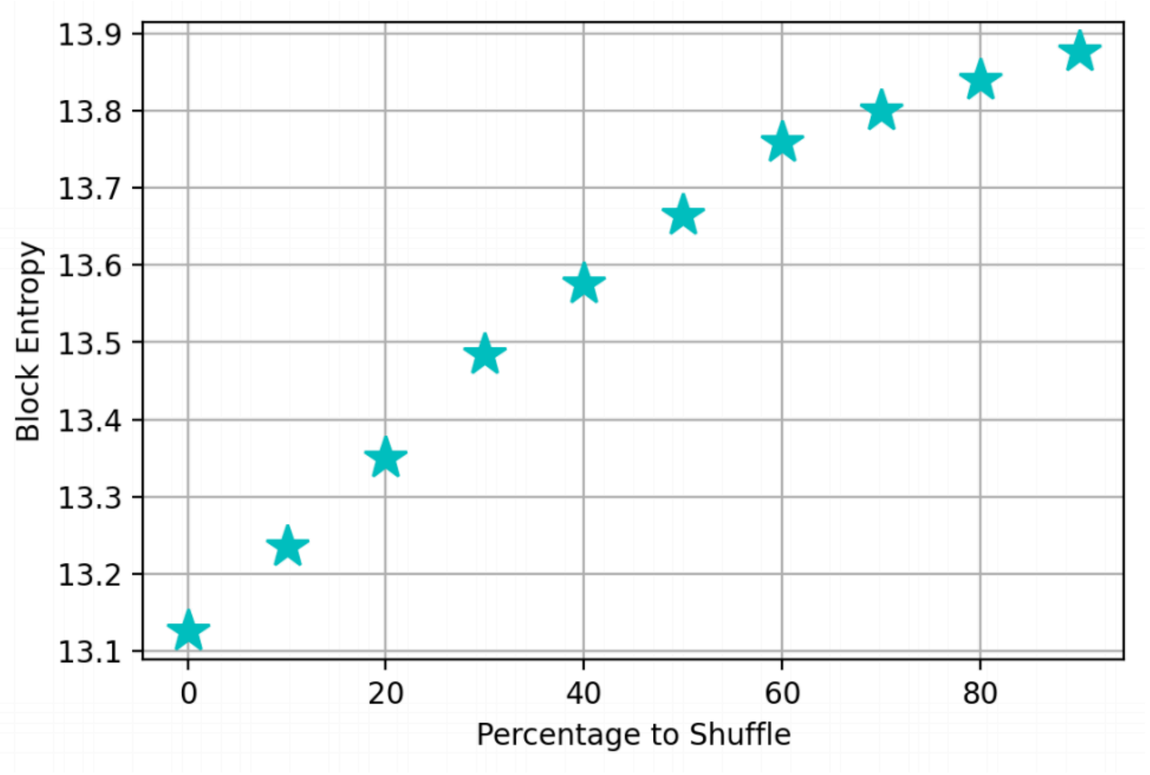
$$\hat{y}_i = \text{argmax}_c \ \hat{y}_{i,c}$$

# Time-Aware MIL Pooling

- 기존 MIL의 Pooling 방식은 순서 불변성

  Mean({A, B, C}) == Mean({C, B, A})

- 시계열 내의 시점들을 섞는 것은 예측 가능성을 방해




SEOULTECH

# Time-Aware MIL Pooling

- 시간적 상관관계(c)를 고려하면 복잡성이 낮아진다

$$H(\boldsymbol{X}_c|\boldsymbol{Y}) \leq H(\boldsymbol{X}_{nc}|\boldsymbol{Y})$$

$$H(\boldsymbol{X}_c|\boldsymbol{Y}) = H((\Theta^1|t^1), \cdots, (\Theta^T|t^T)|\boldsymbol{Y})$$

denote $(\Theta^j|t^j)$ by $\Lambda^j$.

$$\begin{aligned} H(\Lambda^1, \cdots, \Lambda^T|\boldsymbol{Y}) = & H(\Lambda^T|\Lambda^{T-1}, \cdots, \Lambda^1, \boldsymbol{Y}) \\ & + H(\Lambda^{T-1}|\Lambda^{T-2}, \cdots, \Lambda^1, \boldsymbol{Y}) \\ & + \cdots \\ & + H(\Lambda^1|\boldsymbol{Y}) \end{aligned}$$

$$H(\Lambda^1, \cdots, \Lambda^T|\boldsymbol{Y}) \leq \sum_i H(\Lambda^i|\boldsymbol{Y})$$

SEOULTECH

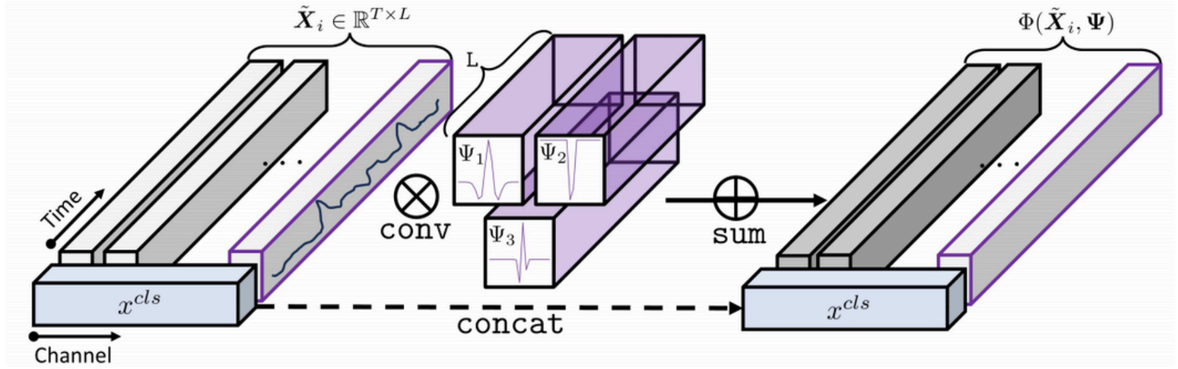
# Time-Aware MIL Pooling

- TimeMIL은 두 가지 방법을 활용해 구현

  1. Wavelet Positional Encoding (순서 문제를 해결)

  2. Self Attention (시간 상관관계를 효율적으로 구현)

# Wavelet Positional Encoding




$$\mathrm{WPE}(\boldsymbol{X}_i) = \sum_{j=1}^{n_\mathrm{W}} \Phi(\boldsymbol{X}_i, \boldsymbol{\Psi}_j)$$

$$\Phi(\tilde{\boldsymbol{X}}_i, \boldsymbol{\Psi}_j) = \begin{bmatrix} \tilde{\boldsymbol{X}}_{i1} \circledast \psi_{a_{j1}, b_{j1}} \\ \vdots \\ \tilde{\boldsymbol{X}}_{iL} \circledast \psi_{a_{jL}, b_{jL}} \end{bmatrix}^\top \in \mathbb{R}^{n \times L}$$

- 기존 Sinusouidal positional Encoding 방법은 사전에 정의된 고정 함수를 사용

$$PE_{(pos,2i)} = \sin(pos/10000^{2i/d_{model}})$$

$$PE_{(pos,2i+1)} = \cos(pos/10000^{2i/d_{model}})$$

- 입력 신호와 웨이블릿 필터 각각 채널별 Conv 연산을 수행
- 최적의 스케일(a)과 위치(b) 파라미터를 학습 – 시간과 주파수 해상도 모두 고려

SEOULTECH

# Wavelet Positional Encoding

- Wavelet Transform

(1) : mother wavelet

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{x-b}{a}\right)$$

(2) : orthogonal

$$\langle \psi_{a,b}, \psi_{a',b'} \rangle \equiv \int_{-\infty}^{\infty} \psi_{a,b}(t)\psi_{a',b'}(t)dt = 0$$

(2) : CWT(Continuous Wavelet Transform)

$$f(a,b) = \int_{-\infty}^{\infty} f(t)\psi_{a,b}(t)dt = f(t) \circledast \psi_{a,b}(t)$$

SEOULTECH

# Self Attention

- 과거의 모든 정보를 조건으로 가지는 것은 비효율적
- 클래스 토큰을 활용해 효율적으로 구현
- Self-Attention을 통해 클래스 토큰이 입력 데이터의(bag) 정보를 요약 및 대표

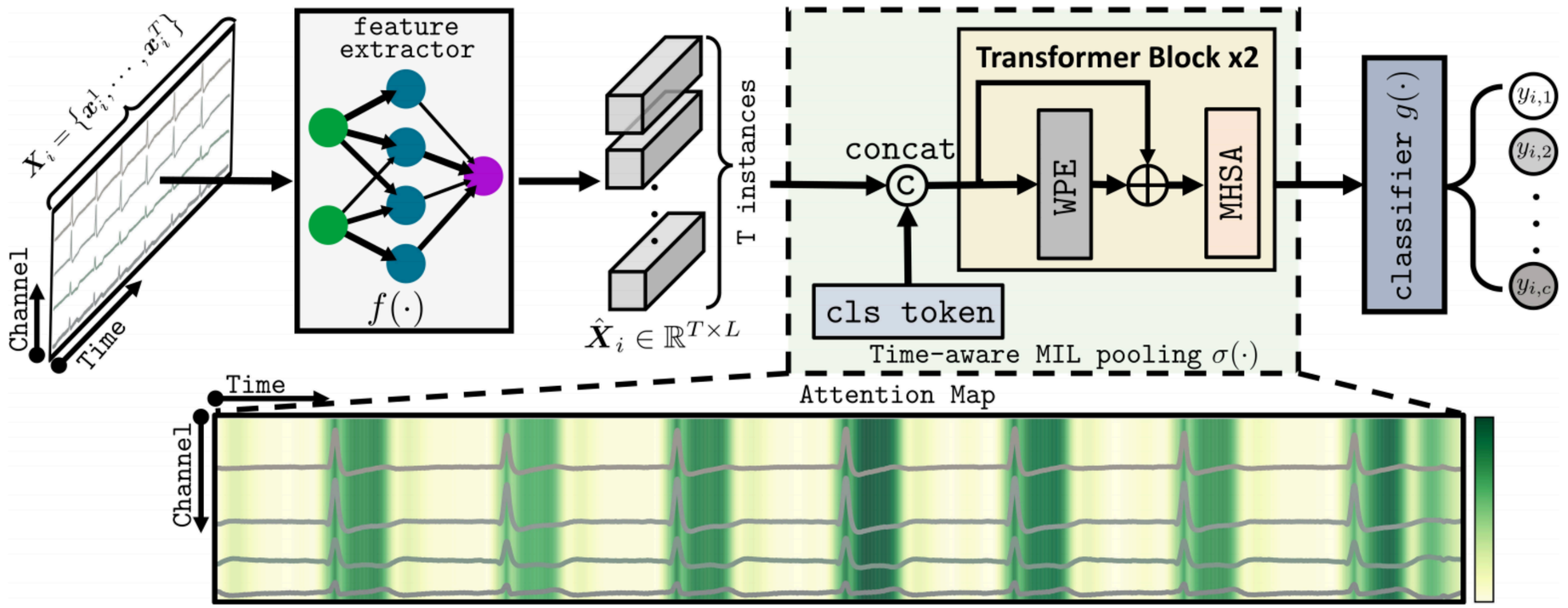
$$\mathrm{MHSA}(\boldsymbol{X}_i^{\mathrm{cls}}) = [\mathrm{head}_1, \cdots, \mathrm{head}_H]\boldsymbol{W}_0$$

$$\mathrm{head}_h = \mathrm{Attention}(\boldsymbol{X}_i^{\mathrm{cls}}\boldsymbol{W}_h^Q, \boldsymbol{X}_i^{\mathrm{cls}}\boldsymbol{W}_h^K, \boldsymbol{X}_i^{\mathrm{cls}}\boldsymbol{W}_h^V)$$

$$\mathrm{Attention}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \mathrm{softmax}(\boldsymbol{Q}\boldsymbol{K}^{\mathbf{T}}/\sqrt{d_k})\boldsymbol{V}$$

# TimeMIL

- TimeMIL은 특징 추출기로 Inception Time 모델 사용



# Result

| Datasets/Methods | ED-1NN | DTW-1NN-I | DTW-1NN-D | MLSTM-FCN *Neur. Net.'19* | ShapeNet *AAAI'21* | WEASEL+MUSE *arxiv'2017* | TapNet *AAAI'20* | OS-CNN *ICLR'22* | MOS-CNN *ICLR'22* | TodyNet *arxiv'23* | Ours |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ArticularyWordRecognition | 0.970 | 0.980 | 0.987 | 0.973 | 0.987 | 0.990 | 0.987 | 0.988 | **0.991** | 0.987 | 0.990 |
| AtrialFibrillation | 0.267 | 0.267 | 0.200 | 0.267 | 0.400 | 0.333 | 0.333 | 0.233 | 0.183 | 0.467 | **0.733** |
| BasicMotions | 0.675 | **1.000** | 0.975 | 0.950 | **1.000** | **1.000** | **1.000** | **1.000** | **1.000** | **1.000** | **1.000** |
| Cricket | 0.944 | 0.986 | **1.000** | 0.917 | 0.986 | **1.000** | 0.958 | 0.993 | 0.990 | **1.000** | **1.000** |
| DuckDuckGeese | 0.275 | 0.550 | 0.600 | 0.675 | 0.725 | 0.575 | 0.575 | 0.540 | 0.615 | 0.580 | **0.780** |
| EigenWorms | 0.550 | 0.603 | 0.618 | 0.504 | 0.878 | **0.890** | 0.489 | 0.414 | 0.508 | 0.840 | 0.823 |
| Epilepsy | 0.667 | 0.978 | 0.964 | 0.761 | 0.987 | **1.000** | 0.971 | 0.980 | 0.996 | 0.971 | **1.000** |
| EthanolConcentration | 0.293 | 0.304 | 0.323 | 0.373 | 0.312 | 0.133 | 0.323 | 0.240 | **0.415** | 0.350 | 0.407 |
| ERing | 0.133 | 0.133 | 0.133 | 0.133 | 0.133 | 0.430 | 0.133 | 0.881 | 0.915 | 0.915 | **0.956** |
| FaceDetection | 0.519 | 0.513 | 0.529 | 0.545 | 0.602 | 0.545 | 0.556 | 0.575 | 0.597 | 0.627 | **0.698** |
| FingerMovements | 0.550 | 0.520 | 0.530 | 0.580 | 0.589 | 0.490 | 0.530 | 0.568 | 0.568 | 0.570 | **0.670** |
| HandMovementDirection | 0.279 | 0.306 | 0.231 | 0.365 | 0.338 | 0.365 | 0.378 | 0.443 | 0.361 | **0.649** | 0.487 |
| Handwriting | 0.371 | 0.509 | 0.607 | 0.286 | 0.451 | 0.605 | 0.357 | 0.668 | **0.677** | 0.436 | 0.482 |
| Heartbeat | 0.620 | 0.659 | 0.717 | 0.663 | 0.756 | 0.727 | 0.751 | 0.489 | 0.604 | 0.756 | **0.815** |
| Libras | 0.833 | 0.894 | 0.872 | 0.856 | 0.856 | 0.878 | 0.850 | 0.950 | 0.965 | 0.850 | **0.972** |
| LSST | 0.456 | 0.575 | 0.551 | 0.373 | 0.590 | 0.590 | 0.568 | 0.413 | 0.521 | 0.615 | **0.690** |
| MotorImagery | 0.510 | 0.390 | 0.500 | 0.510 | 0.610 | 0.500 | 0.590 | 0.535 | 0.515 | 0.640 | **0.720** |
| NATOPS | 0.860 | 0.850 | 0.883 | 0.889 | 0.883 | 0.870 | 0.939 | 0.968 | 0.951 | 0.972 | **0.994** |
| PenDigits | 0.973 | 0.939 | 0.977 | 0.978 | 0.977 | 0.948 | 0.980 | 0.985 | 0.983 | **0.987** | 0.600 |
| PEMS-SF | 0.705 | 0.734 | 0.711 | 0.699 | 0.751 | N/A | 0.751 | 0.760 | 0.764 | 0.780 | **0.931** |
| PhonemeSpectra | 0.104 | 0.151 | 0.151 | 0.110 | 0.298 | 0.190 | 0.175 | 0.299 | 0.295 | 0.309 | **0.311** |
| RacketSports | 0.868 | 0.842 | 0.803 | 0.803 | 0.882 | **0.934** | 0.868 | 0.877 | 0.929 | 0.803 | 0.908 |
| SelfRegulationSCP1 | 0.771 | 0.765 | 0.775 | 0.874 | 0.782 | 0.710 | 0.652 | 0.835 | 0.829 | **0.898** | **0.898** |
| SelfRegulationSCP2 | 0.483 | 0.533 | 0.539 | 0.472 | 0.578 | 0.460 | 0.550 | 0.532 | 0.510 | 0.550 | **0.639** |
| StandWalkJump | 0.200 | 0.333 | 0.200 | 0.067 | 0.533 | 0.333 | 0.400 | 0.383 | 0.383 | 0.467 | **0.733** |
| UWaveGestureLibrary | 0.881 | 0.869 | 0.903 | 0.891 | 0.906 | 0.916 | 0.894 | **0.927** | 0.926 | 0.850 | 0.900 |
| Ours 1-to-1-Wins | 25 | 23 | 22 | 25 | 22 | 16 | 24 | 22 | 19 | 20 | - |
| Ours 1-to-1-Draws | 0 | 1 | 1 | 0 | 1 | 4 | 1 | 1 | 1 | 3 | - |
| Ours 1-to-1-Losses | 1 | 2 | 3 | 1 | 3 | 5 | 1 | 3 | 6 | 3 | - |
| Average accuracy (↑) | 0.568 | 0.622 | 0.626 | 0.597 | 0.684 | 0.656 | 0.637 | 0.672 | 0.692 | 0.726 | **0.774** |
| Total best accuracy (↑) | 0 | 1 | 1 | 0 | 1 | 5 | 1 | 2 | 4 | 5 | **18** |
| Average Rank (↓) | 9.154 | 7.904 | 7.231 | 7.865 | 4.731 | 5.900 | 6.500 | 5.442 | 4.692 | 4.058 | **2.327** |

SEOULTECH

# Result

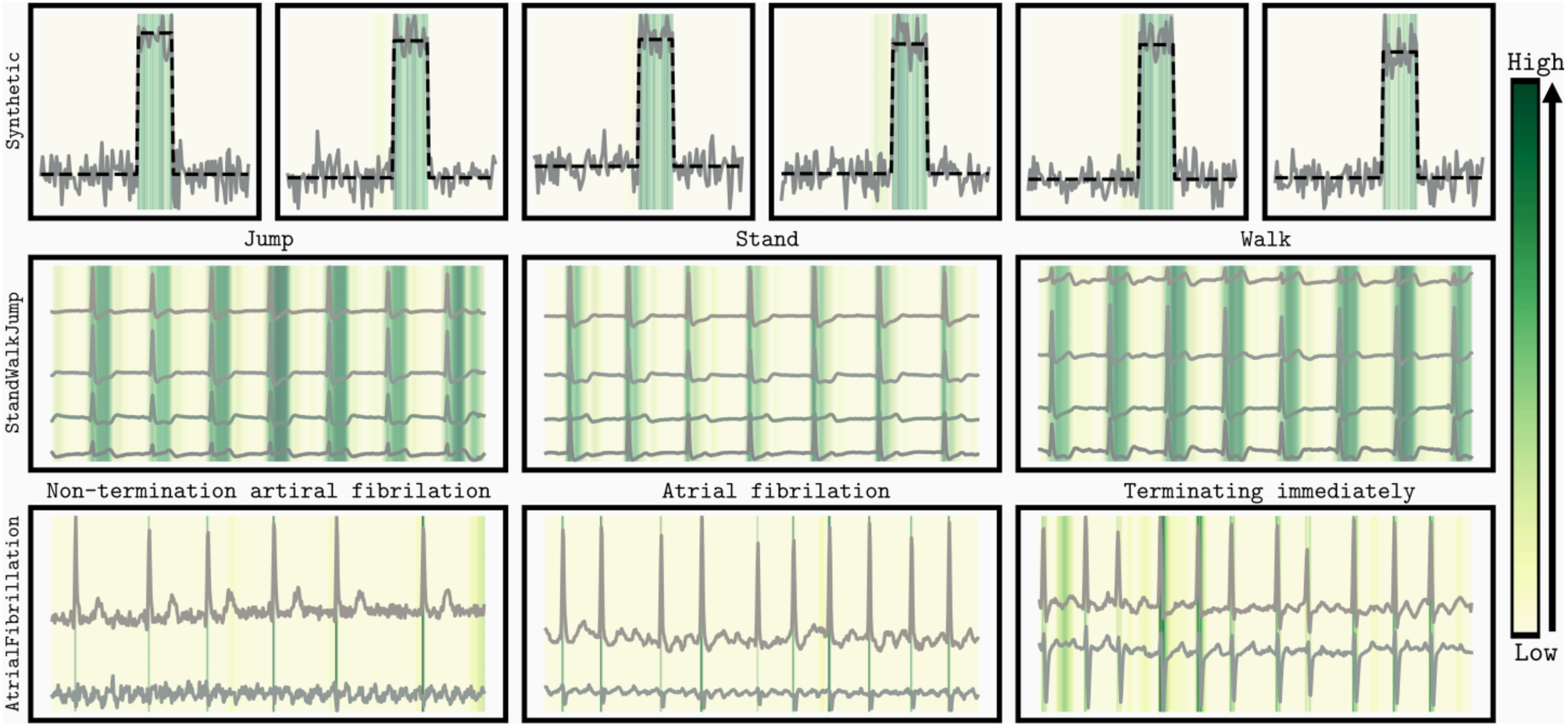
| Datasets/Methods | LSTNet *SIGIR'18* | LSSL *ICLR'22* | Rocket *DMKD'18* | SCINet *NeurIPS'22* | MICN *ICLR'23* | TimesNet *ICLR'23* | Dlinear *AAAI'23* | LightTS *arxiv'22* | MTS-Mixer *arxiv'23* | Rlinear *arxiv'23* | RMLP *arxiv'23* | FEDformer *ICML'22* | Flowformer *ICML'22* | Crossformer *ICLR'23* | PatchTST *ICLR'23* | ModernTCN *ICLR'24* | Ours |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| EthanolConcentration | 0.399 | 0.311 | **0.452** | 0.344 | 0.353 | 0.357 | 0.362 | 0.297 | 0.338 | 0.289 | 0.313 | 0.312 | 0.338 | 0.380 | 0.328 | 0.363 | 0.407 |
| FaceDetection | 0.657 | 0.667 | 0.647 | 0.689 | 0.652 | 0.686 | 0.680 | 0.675 | 0.702 | 0.656 | 0.673 | 0.660 | 0.676 | 0.687 | 0.683 | **0.708** | 0.698 |
| Handwriting | 0.258 | 0.246 | **0.588** | 0.236 | 0.255 | 0.321 | 0.270 | 0.261 | 0.260 | 0.281 | 0.300 | 0.280 | 0.338 | 0.288 | 0.296 | 0.306 | 0.487 |
| Heartbeat | 0.771 | 0.727 | 0.756 | 0.775 | 0.747 | 0.780 | 0.751 | 0.751 | 0.771 | 0.726 | 0.727 | 0.737 | 0.776 | 0.776 | 0.749 | 0.772 | **0.815** |
| JapaneseVowels | 0.981 | 0.984 | 0.962 | 0.960 | 0.946 | 0.984 | 0.962 | 0.962 | 0.943 | 0.959 | 0.959 | 0.984 | 0.989 | 0.991 | 0.975 | 0.988 | **0.995** |
| PEMS-SF | 0.867 | 0.861 | 0.751 | 0.838 | 0.855 | 0.896 | 0.751 | 0.884 | 0.809 | 0.827 | 0.839 | 0.809 | 0.860 | 0.859 | 0.893 | 0.891 | **0.931** |
| SelfRegulationSCP1 | 0.840 | 0.908 | 0.908 | 0.925 | 0.860 | 0.918 | 0.873 | 0.898 | 0.917 | 0.911 | 0.921 | 0.887 | 0.925 | 0.921 | 0.907 | **0.934** | 0.898 |
| SelfRegulationSCP2 | 0.528 | 0.522 | 0.533 | 0.572 | 0.536 | 0.572 | 0.505 | 0.511 | 0.550 | 0.561 | 0.510 | 0.544 | 0.561 | 0.583 | 0.578 | 0.603 | **0.639** |
| SpokenArabicDigits | **1.000** | **1.000** | 0.712 | 0.981 | 0.971 | 0.990 | 0.814 | **1.000** | 0.974 | 0.965 | 0.976 | **1.000** | 0.988 | 0.979 | 0.983 | 0.987 | **1.000** |
| UWaveGestureLibrary | 0.878 | 0.859 | **0.944** | 0.851 | 0.828 | 0.853 | 0.821 | 0.803 | 0.823 | 0.825 | 0.838 | 0.853 | 0.866 | 0.853 | 0.858 | 0.867 | 0.900 |
| Ours 1-to-1-Wins | 9 | 8 | 6 | 9 | 10 | 9 | 10 | 8 | 8 | 9 | 9 | 9 | 9 | 9 | 9 | 8 | - |
| Ours 1-to-1-Draws | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | - |
| Ours 1-to-1-Losses | 0 | 1 | 4 | 1 | 0 | 1 | 0 | 0 | 2 | 1 | 1 | 0 | 1 | 1 | 1 | 2 | - |
| Average accuracy (↑) | 0.718 | 0.709 | 0.725 | 0.717 | 0.700 | 0.736 | 0.679 | 0.704 | 0.709 | 0.700 | 0.706 | 0.707 | 0.732 | 0.732 | 0.725 | 0.742 | **0.777** |
| Total best accuracy (↑) | 1 | 1 | 3 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 2 | **5** |
| Average Rank (↓) | 8.85 | 10.4 | 9.5 | 8.9 | 13.2 | 5.35 | 12.7 | 11.2 | 10.85 | 13 | 11.55 | 10.75 | 5.9 | 5.9 | 8.1 | 4 | **2.85** |

| MIL Pooling | Accuracy |
|---|---|
| Mean | 0.715 |
| Max | 0.719 |
| Attention | 0.739 |
| Conjunctive | 0.746 |
| **Time-aware (Ours)** | **0.774** |

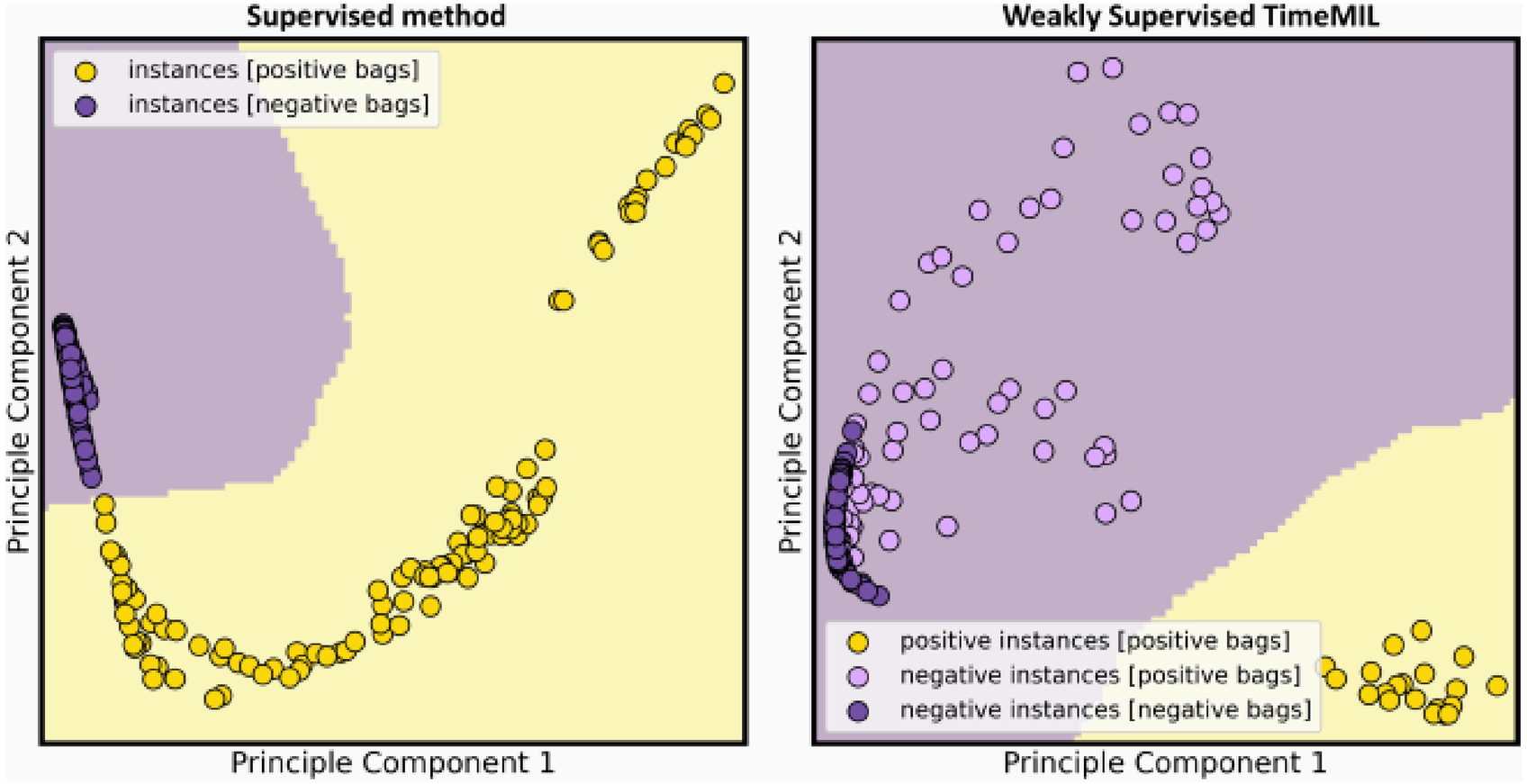
| Positional Encoding | Accuracy | |
|---|---|---|
| | ABMIL | TimeMIL |
| None | 0.739 | 0.761 |
| Sinusoidal | 0.741 | 0.763 |
| **WPE (Ours)** | 0.745 | **0.774** |

SEOULTECH

# Result



# Result




SEOULTECH

# Conclusion

- TimeMIL은 28개의 데이터셋에서 26개의 방법보다 우월한 성능을 입증
- 시간적 종속성과 순서에 초점을 맞추고 있어 해석 가능한 폴링을 채널 수준으로 확장하지 못함
- 다변량을 융합하는 경우에는 한계가 존재

SEOULTECH