NOVEL DEEP LEARNING MODELS AND TECHNIQUES FOR EFFECTIVE AND EFFICIENT HANDLING OF MULTIPLE DATA MODALITIES

Qiang Shawn Cheng

Institute for Biomedical Informatics

Department of Internal Medicine

Department of Computer Science

University of Kentucky

BRIEF OUTLINE

Four Topics

- A New Technique for Tabular Data Learning Problems
- A New Deep Learning Model for Long-Term Time Series Forecasting
- A New Diversity-Aware Tensor-Formulated Multi-View Clustering Technique
- A New Deep Learning Model for Time Series Classification
- (Optional) An Unsupervised Circadian Phase Inference Technique for Multi-Omics Data with Applications to Neurodegenerative Diseases

ACKNOWLEDGEMENT AND DISCLAIMER

The presented contents are based on joint work with Md Atik Ahamed, Aram Ansary Ogholbake, Dr. Peng, et al.

The research is partially supported by NSF under Grant IIS 2327113 and ITE 2433190, and the NIH under Grants R21AG070909, P30AG072946, and R01HD101508-01. I declare that the results presented in this talk do not pose any potential conflicts of interest with any funding agencies or collaborators involved in this research. The findings and conclusions are based on impartial and objective analysis, and no external influences have compromised the integrity of this work.

1ST TOPIC: A NEW DEEP LEARNING APPROACH FOR HANDLING TABULAR DATA

Motivation

- Importance of tabular data across domains
  - Structured format; mixed types; predominant in various domains
- Challenges with existing deep learning models for tabular data
  - Classical methods (MLP, XGBoost, LR): accuracy
  - Deep learning models (TabNet, CNN-based, DCN): accuracy
  - Transformer-based models (AutoInt, TabTransformer, FT-former, TransTab): efficiency, accessibility, and scalability
  - Self-Supervised Learning models (VIME, SCARF): accuracy
- Need for an accurate, efficient, scalable approach
  - Capturing correlations across features and samples
  - Also, existing models not suited to feature incremental learning

NEW MODEL: MAMBATAB

- Novel approach leveraging structured state-space model (SSM)

$$dh(t)/dt = A\,h(t) + B\,u(t), \quad x(t) = C\,h(t),$$

- Based on Mamba, an SSM variant

$$h_k = \bar{A}\,h_{k-1} + \bar{B}\,u_k, \quad x_k = C\,h_k, \qquad \bar{A} = \exp(\Delta A), \quad \bar{B} = (\Delta A)^{-1}(\exp(\Delta A) - I)\Delta B.$$

  - B, C, Δ: time varying (dependent on input); A: diagonal

- Advantages:
  - Small model size and number of parameters
  - Minimal preprocessing
  - High performance in accuracy
  - Linear complexity and linear scalability

MambaTab: A plug-and-play model for learning tabular data. MA Ahamed, Q Cheng, IEEE MIPR 2024.
Code available at GitHub.

METHOD - DATA PREPROCESSING

Simple, minimal preprocessing

- Normalization and imputation
  - Min-max scaling

METHOD - EMBEDDING REPRESENTATION LEARNING

- Encoding categorical/Binary variables into numerical var. automatically
  - Each sample encoded into a token in a latent space

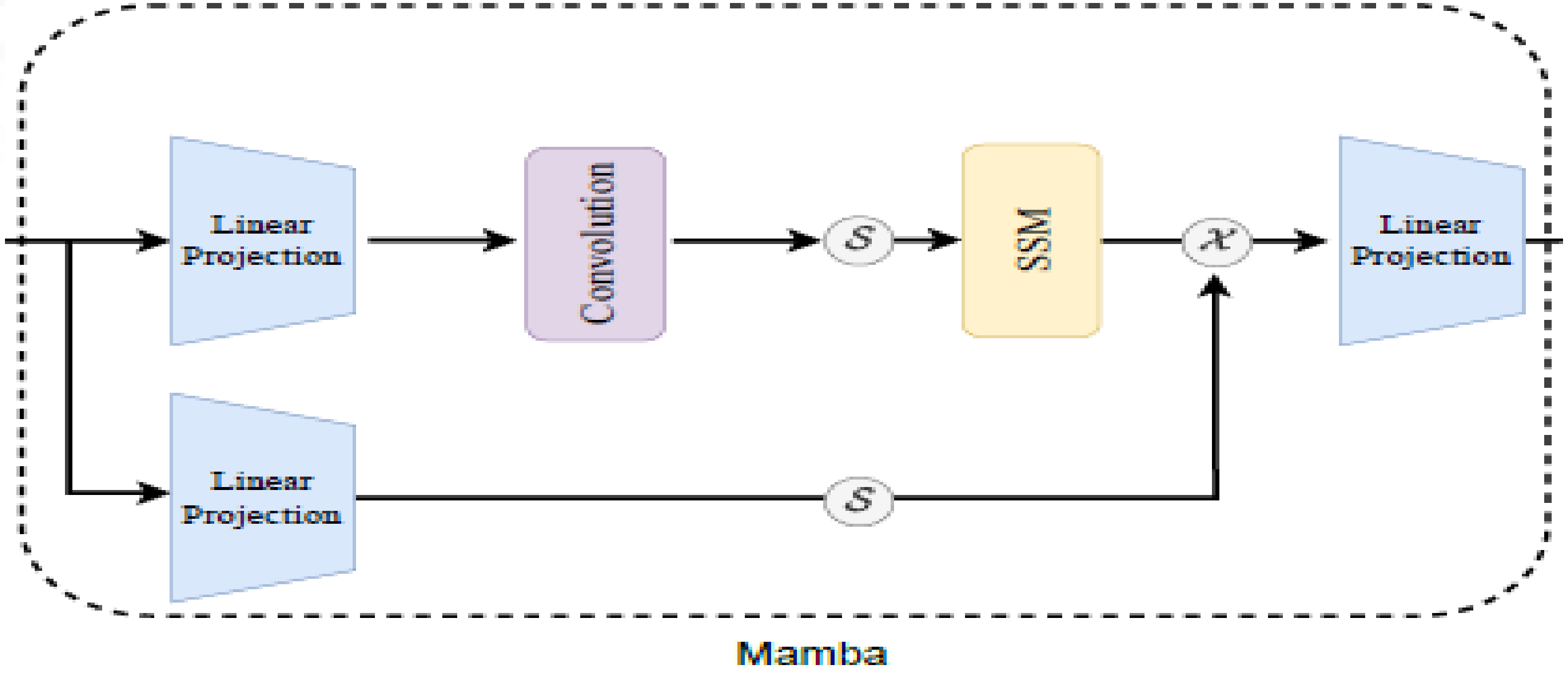
- Fully connected layer to learn embedded representations

- Enabling meaningful representations and incremental learning

- Layer normalization applied

METHOD - CASCADING MAMBA BLOCKS

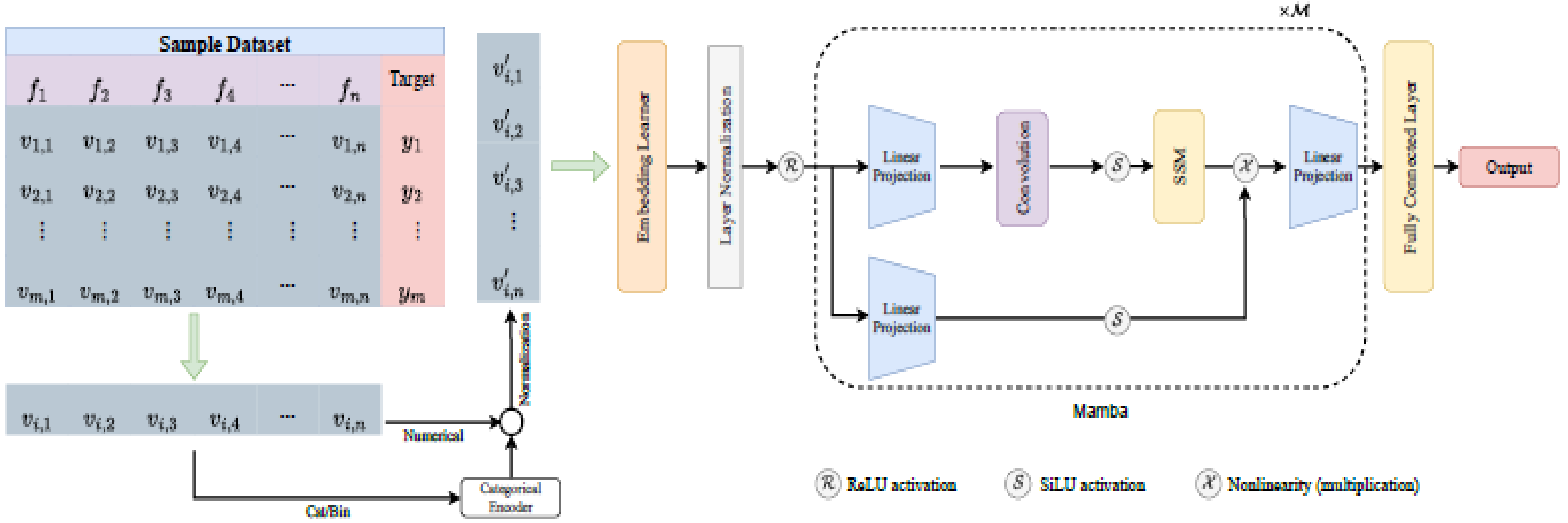
- Mamba block maps features into a feature space of the same dim.
- Utilizes linear projections, 1D causal convolution, SiLU activation, SSMs



- Context-dependent feature extraction and reasoning for long-range dependencies
- Cascading multiple Mamba blocks to seek important information w. various contexts

METHOD - OUTPUT PREDICTION

- Fully connected layer maps Mamba block outputs to predictions
- Sigmoid activation for probability scores



EXPERIMENTAL SETUP

- Datasets: 8 diverse public tabular datasets

Dataset Name	Abbreviation	Datapoints	Train	Val	Test	Positive
Credit-g	CG	1000	700	100	200	0.70
Credit-approval	CA	690	483	69	138	0.56
Dresses-sales	DS	500	350	50	100	0.42
Adult	AD	48842	34189	4884	9769	0.24
Cylinder-bands	CB	540	378	54	108	0.58
Blastchar	BL	7043	4930	704	1409	0.27
Insurance-co	IO	5822	4075	582	1165	0.06
Income-1995	IC	32561	22792	3256	6513	0.24

- Baselines: 12 SOTA models - LR, XGBoost, MLP, SNN, TabNet, DCN, AutoInt, TabTransformer, FT-Transformer, VIME, SCARF, TransTab
- Settings: Vanilla supervised learning, feature incremental learning
- Metrics: AUROC

RESULTS - VANILLA SUPERVISED LEARNING

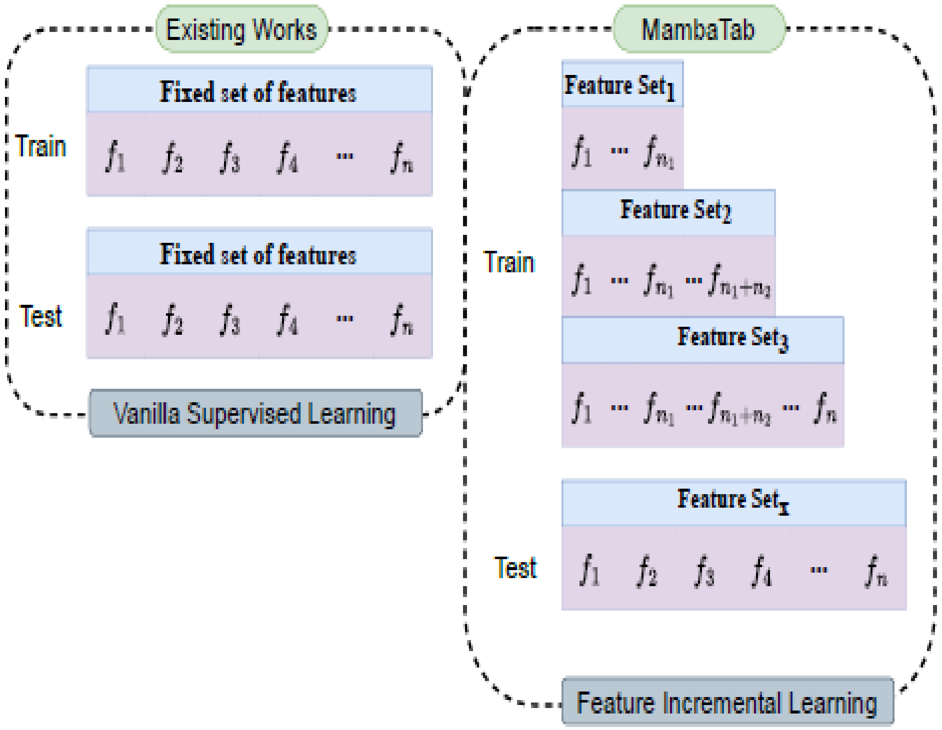
AUROC results on 8 datasets (averaged over 10 runs)

Methods	Datasets							
	CG	CA	DS	AD	CB	BL	IO	IC
LR	0.720	0.836	0.557	0.851	0.748	0.801	0.769	0.860
XGBoost	0.726	0.895	0.587	0.912	0.892	0.821	0.758	**0.925**
MLP	0.643	0.832	0.568	0.904	0.613	0.832	0.779	0.893
SNN	0.641	0.880	0.540	0.902	0.621	0.834	0.794	0.892
TabNet	0.585	0.800	0.478	0.904	0.680	0.819	0.742	0.896
DCN	0.739	0.870	0.674	0.913	0.848	0.840	0.768	0.915
AutoInt	0.744	0.866	0.672	0.913	0.808	0.844	0.762	0.916
TabTrans	0.718	0.860	0.648	**0.914**	0.855	0.820	0.794	0.882
FT-Trans	0.739	0.859	0.657	0.913	0.862	0.841	0.793	0.915
VIME	0.735	0.852	0.485	0.912	0.769	0.837	0.786	0.908
SCARF	0.733	0.861	0.663	0.911	0.719	0.833	0.758	0.905
TransTab	0.768	0.881	0.643	0.907	0.851	0.845	**0.822**	0.919
MambaTab-D	0.771	0.954	0.643	0.906	0.862	0.852	0.785	0.906
MambaTab-T	**0.801**	**0.963**	**0.681**	**0.914**	**0.896**	**0.854**	0.812	0.920

- MambaTab outperforms baselines on majority datasets
- Hyperparameter tuning (MambaTab-T) improves upon default model

University of Kentucky

RESULTS - FEATURE INCREMENTAL LEARNING



AUROC results under feature incremental setting

Methods	Datasets							
	CG	CA	DS	AD	CB	BL	IO	IC
LR	0.670	0.773	0.475	0.832	0.727	0.806	0.655	0.825
XGBoost	0.608	0.817	0.527	0.891	0.778	0.816	0.692	0.898
MLP	0.586	0.676	0.516	0.890	0.631	0.825	0.626	0.885
SNN	0.583	0.738	0.442	0.888	0.644	0.818	0.643	0.881
TabNet	0.573	0.689	0.419	0.886	0.571	0.837	0.680	0.882
DCN	0.674	0.835	0.578	0.893	0.778	0.840	0.660	0.891
AutoInt	0.671	0.825	0.563	0.893	0.769	0.836	0.676	0.887
TabTrans	0.653	0.732	0.584	0.856	0.784	0.792	0.674	0.828
FT-Trans	0.662	0.824	0.626	0.892	0.768	0.840	0.645	0.889
VIME	0.621	0.697	0.571	0.892	0.769	0.803	0.683	0.881
SCARF	0.651	0.753	0.556	0.891	0.703	0.829	0.680	0.887
TransTab	0.741	0.879	0.665	0.894	0.791	0.841	0.739	0.897
MambaTab-D	**0.787**	**0.961**	**0.669**	**0.904**	**0.860**	**0.853**	**0.783**	**0.908**

- MambaTab outperforms TransTab and other basdelines with default hyperparameters on all datasets

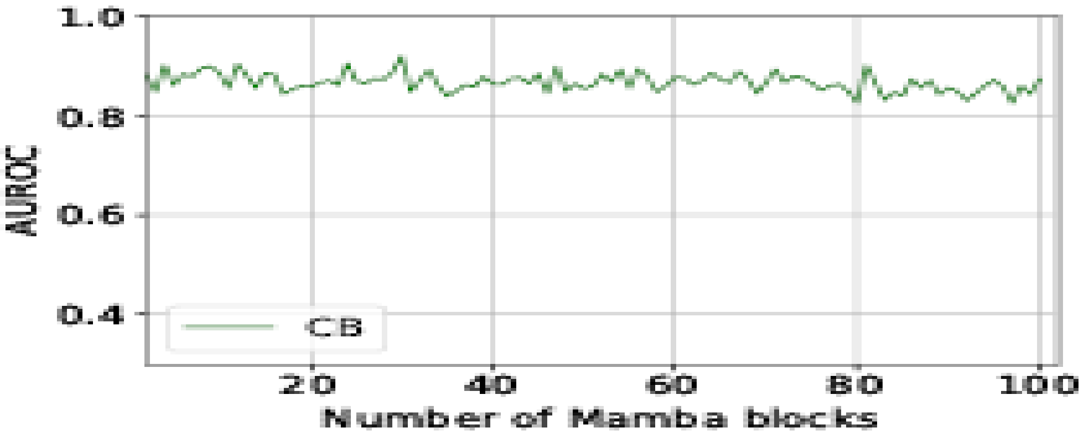
ANALYSIS - PARAMETER EFFICIENCY

Comparing learnable parameters of MambaTab vs Transformers

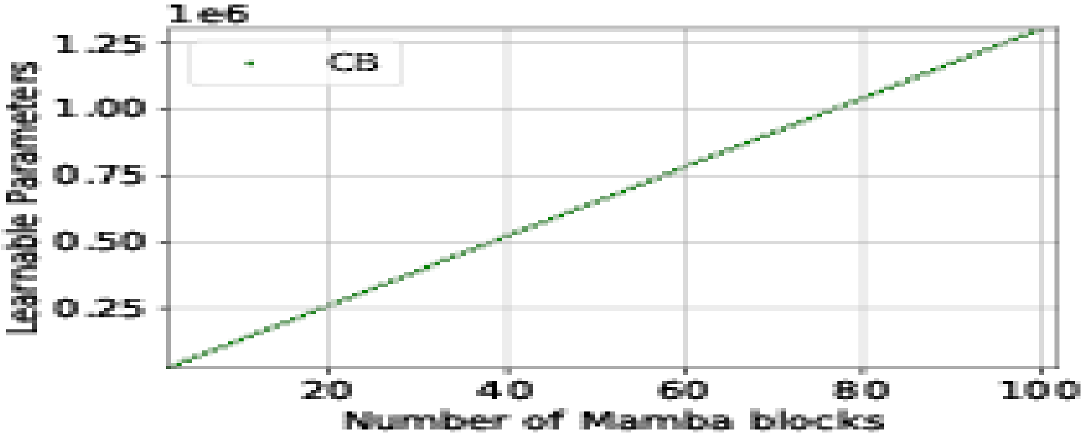
Methods	Datasets							
	CG	CA	DS	AD	CB	BL	IO	IC
TabTrans	2.7M	1.2M	2.0M	1.2M	6.5M	3.4M	87.0M	1.0M
FT-Trans	176K	176K	179K	178K	203K	176K	193K	177K
TransTab	4.2M	4.2M	4.2M	4.2M	4.2M	4.2M	4.2M	4.2M
MambaTab-D	**13K**	**13K**	13K	**13K**	**14K**	13K	15K	13K
MambaTab-T	50K	38K	**5K**	255K	30K	**11K**	**13K**	**10K**

- MambaTab uses <1% parameters of TransTab while achieving better performance

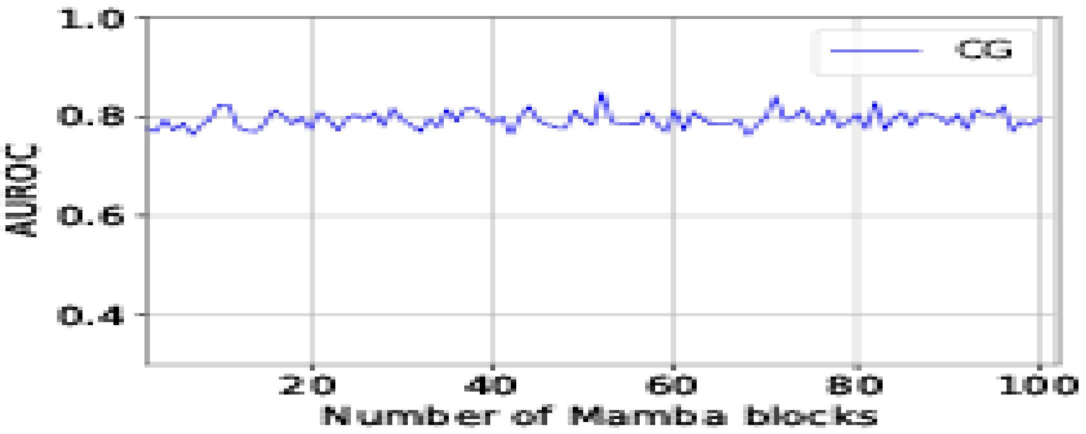
ANALYSIS - SCALABILITY AND ABLATION STUDY



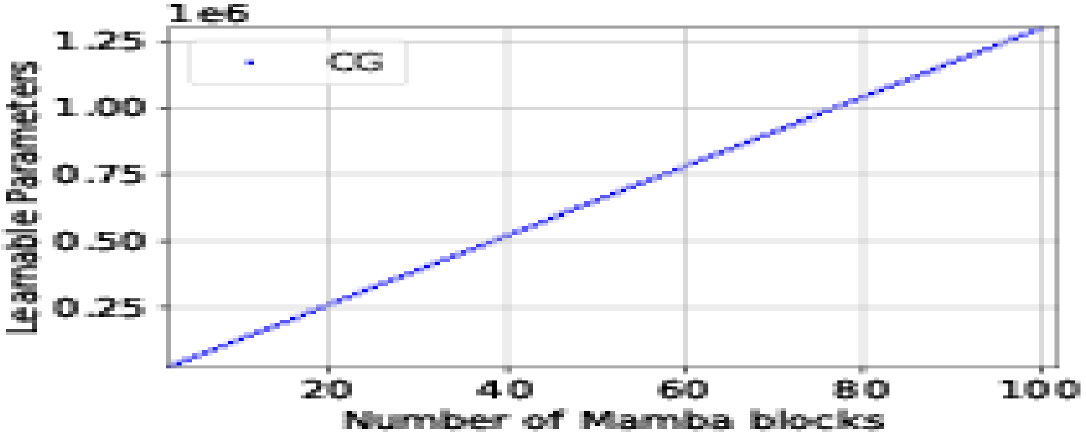
(a) AUROC results (CB)



(b) Learnable parameters (CB)



(c) AUROC results (CG)



(d) Learnable parameters (CG)

- Extensive hyper-parameter sensitivity study and structural ablations study
  - More in the paper – MambaTab - Not shown here

BRIEF CONCLUSION OF THE 1ST TOPIC

- MambaTab: out-of-the-box, plug-and-play model for tabular data
- Simple architecture, minimal preprocessing, superior performance, parameter-efficient
- We also obtained superior results for SSL, not shown here
- Holds promise for enabling wider practical applications across domains

2ND TOPIC – A DEEP LEARNING MODEL FOR TIME SERIES LONG-TERM FORECASTING

- Importance of long-term time-series forecasting (LTSF) across domains
- Challenges in LTSF
  - capturing long-term dependencies
  - scalability
  - computational efficiency
- Limitations of existing approaches
  - Non-Transformer-based: Classical models (ARIMA, RNN, GARCH), Linear models – MLP (DLinear, RLinear, TiDE), CNN (TimesNet, Scinet)
    linear complexity and scalability, may not capture LT dependency, accuracy
  - Transformer-based SL models (iTransformer, PatchTST, CrossFormer): can capture LT dependency, SOTA accuracy, quadratic complexity and not scalable well
  - SSL Representation Learning: typically Transformer-based (TST, TS-TCC):
    not as competitive as SL models in accuracy

NEW MODEL: TIMEMACHINE

- Novel approach leveraging structured state-space models (SSMs), Mamba
- Exploiting unique properties of time series data to produce salient contextual cues at multi-scales
- Innovative integrated quadruple-Mamba architecture

- Key Innovations
  - First to leverage purely SSM modules for context-aware LTSF with linear scalability and small memory footprints
  - Unifies handling of channel-mixing and channel-independence situations
  - Selects contents for prediction incorporating global and local contexts at different scales

TimeMachine: A Time Series is Worth 4 Mambas for Long-term Forecasting. MA Ahamed, Q Cheng, ECAI, 2024. Code Available at GitHub.

METHOD - DATA PREPROCESSING AND EMBEDDED REPRESENTATIONS

Preprocessing

- Normalization options: RevIN or Z-score
- Channel mixing vs. channel independence handling

Embedded Representation

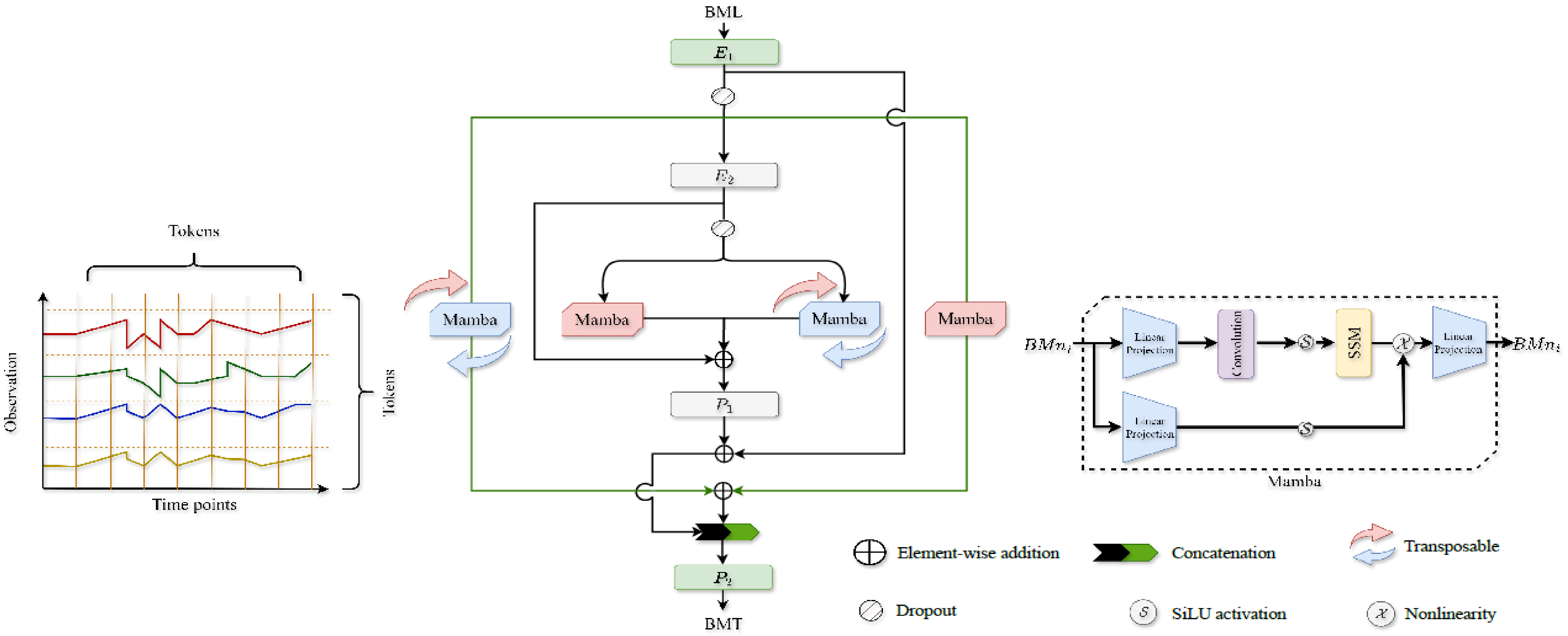
- Two-stage embedded representations using MLPs
- Enables handling variable input sequence lengths

METHOD - INTEGRATED QUADRUPLE MAMBAS

- Two pairs of Mambas at 2 embedding levels
- Capture long-term dependencies and provide local contexts
- Handling channel-mixing and channel-independence cases
- Transposable Mambas for unified architecture

METHOD - OUTPUT PROJECTION AND ARCHITECTURE

- Two-stage output projection using MLPs
- Residual connections for stabilization and overfitting reduction



EXPERIMENTAL SETUP

- Datasets: 7 standard benchmark datasets

Dataset ($\mathcal{D}$)	Channels (M)	Time Points	Frequency
Weather	21	52696	10 Minutes
Traffic	862	17544	Hourly
Electricity	321	26304	Hourly
ETTh1	7	17420	Hourly
ETTh2	7	17420	Hourly
ETTm1	7	69680	15 Minutes
ETTm2	7	69680	15 Minutes

- Baselines: 11 SOTA models
- Metrics: MSE and MAE

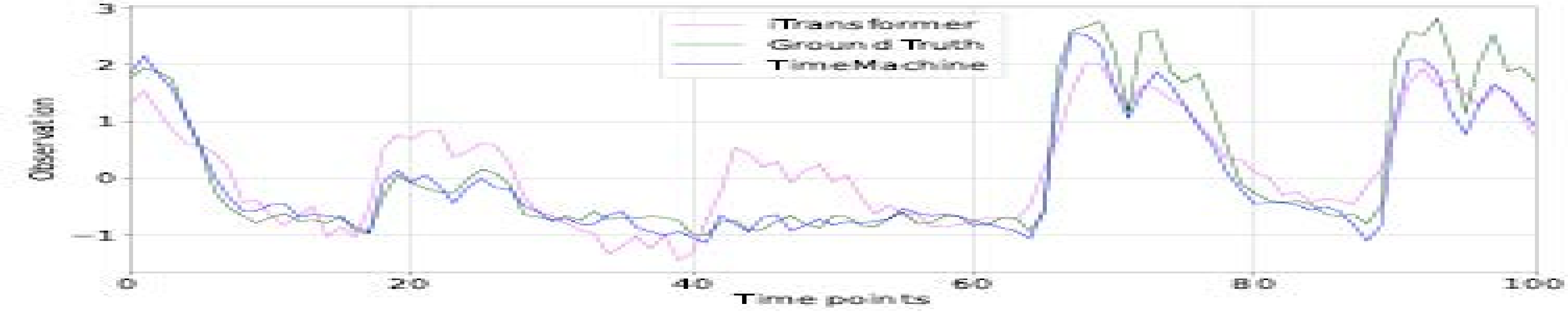
RESULTS - QUANTITATIVE

Methods→		TimeMachine		iTransformer		RLinear		PatchTST		Crossformer		TiDE		TimesNet		DLinear		SCINet		FEDformer		Stationary		Autoformer	
$\mathcal{D}$	T	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Weather	96	0.164	0.208	0.174	0.214	0.192	0.232	0.177	0.218	0.158	0.230	0.202	0.261	0.172	0.220	0.196	0.255	0.221	0.306	0.217	0.296	0.173	0.223	0.266	0.336
	192	0.211	0.250	0.221	0.254	0.240	0.271	0.225	0.259	0.206	0.277	0.242	0.298	0.219	0.261	0.237	0.296	0.261	0.340	0.276	0.336	0.245	0.285	0.307	0.367
	336	0.256	0.290	0.278	0.296	0.292	0.307	0.278	0.297	0.272	0.335	0.287	0.335	0.280	0.306	0.283	0.335	0.309	0.378	0.339	0.380	0.321	0.338	0.359	0.395
	720	0.342	0.343	0.358	0.349	0.364	0.353	0.354	0.348	0.398	0.418	0.351	0.386	0.365	0.359	0.345	0.381	0.377	0.427	0.403	0.428	0.414	0.410	0.419	0.428
Traffic	96	0.397	0.268	0.395	0.268	0.649	0.389	0.544	0.359	0.522	0.290	0.805	0.493	0.593	0.321	0.650	0.396	0.788	0.499	0.587	0.366	0.612	0.338	0.613	0.388
	192	0.417	0.274	0.417	0.276	0.601	0.366	0.540	0.354	0.530	0.293	0.756	0.474	0.617	0.336	0.598	0.370	0.789	0.505	0.604	0.373	0.613	0.340	0.616	0.382
	336	0.433	0.281	0.433	0.283	0.609	0.369	0.551	0.358	0.558	0.305	0.762	0.477	0.629	0.336	0.605	0.373	0.797	0.508	0.621	0.383	0.618	0.328	0.622	0.337
	720	0.467	0.300	0.467	0.302	0.647	0.387	0.586	0.375	0.589	0.328	0.719	0.449	0.640	0.350	0.645	0.394	0.841	0.523	0.626	0.382	0.653	0.355	0.660	0.408
Electricity	96	0.142	0.236	0.148	0.240	0.201	0.281	0.195	0.285	0.219	0.314	0.237	0.329	0.168	0.272	0.197	0.282	0.247	0.345	0.193	0.308	0.169	0.273	0.201	0.317
	192	0.158	0.250	0.162	0.253	0.201	0.283	0.199	0.289	0.231	0.322	0.236	0.330	0.184	0.289	0.196	0.285	0.257	0.355	0.201	0.315	0.182	0.286	0.222	0.334
	336	0.172	0.268	0.178	0.269	0.215	0.298	0.215	0.305	0.246	0.337	0.249	0.344	0.198	0.300	0.209	0.301	0.269	0.369	0.214	0.329	0.200	0.304	0.231	0.338
	720	0.207	0.298	0.225	0.317	0.257	0.331	0.256	0.337	0.280	0.363	0.284	0.373	0.220	0.320	0.245	0.333	0.299	0.390	0.246	0.355	0.222	0.321	0.254	0.361
ETTh1	96	0.364	0.387	0.386	0.405	0.386	0.395	0.414	0.419	0.423	0.448	0.479	0.464	0.384	0.402	0.386	0.400	0.654	0.599	0.376	0.419	0.513	0.491	0.449	0.459
	192	0.415	0.416	0.441	0.436	0.437	0.424	0.460	0.445	0.471	0.474	0.525	0.492	0.436	0.429	0.437	0.432	0.719	0.631	0.420	0.448	0.534	0.504	0.500	0.482
	336	0.429	0.421	0.487	0.458	0.479	0.446	0.501	0.466	0.570	0.546	0.565	0.515	0.491	0.469	0.481	0.459	0.778	0.659	0.459	0.465	0.588	0.535	0.521	0.496
	720	0.458	0.453	0.503	0.491	0.481	0.470	0.500	0.488	0.653	0.621	0.594	0.558	0.521	0.500	0.519	0.516	0.836	0.699	0.506	0.507	0.643	0.616	0.514	0.512
ETTh2	96	0.275	0.334	0.297	0.349	0.288	0.338	0.302	0.348	0.745	0.584	0.400	0.440	0.340	0.374	0.333	0.387	0.707	0.621	0.358	0.397	0.476	0.458	0.346	0.388
	192	0.349	0.381	0.380	0.400	0.374	0.390	0.388	0.400	0.877	0.656	0.528	0.509	0.402	0.414	0.477	0.476	0.860	0.689	0.429	0.439	0.512	0.493	0.456	0.452
	336	0.340	0.381	0.428	0.432	0.415	0.426	0.426	0.433	1.043	0.731	0.643	0.571	0.452	0.452	0.594	0.541	1.000	0.744	0.496	0.487	0.552	0.551	0.482	0.486
	720	0.411	0.433	0.427	0.445	0.420	0.440	0.431	0.446	1.104	0.763	0.874	0.679	0.462	0.468	0.831	0.657	1.249	0.838	0.463	0.474	0.562	0.560	0.515	0.511
ETTm1	96	0.317	0.355	0.334	0.368	0.355	0.376	0.329	0.367	0.404	0.426	0.364	0.387	0.338	0.375	0.345	0.372	0.418	0.438	0.379	0.419	0.386	0.398	0.505	0.475
	192	0.357	0.378	0.377	0.391	0.391	0.392	0.367	0.385	0.450	0.451	0.398	0.404	0.374	0.387	0.380	0.389	0.439	0.450	0.426	0.441	0.459	0.444	0.553	0.496
	336	0.379	0.399	0.426	0.420	0.424	0.415	0.399	0.410	0.532	0.515	0.428	0.425	0.410	0.411	0.413	0.413	0.490	0.485	0.445	0.459	0.495	0.464	0.621	0.537
	720	0.445	0.436	0.491	0.459	0.487	0.450	0.454	0.439	0.666	0.589	0.487	0.461	0.478	0.450	0.474	0.453	0.595	0.550	0.543	0.490	0.585	0.516	0.671	0.561
ETTm2	96	0.175	0.256	0.180	0.264	0.182	0.265	0.175	0.259	0.287	0.366	0.207	0.305	0.187	0.267	0.193	0.292	0.286	0.377	0.203	0.287	0.192	0.274	0.255	0.339
	192	0.239	0.299	0.250	0.309	0.246	0.304	0.241	0.302	0.414	0.492	0.290	0.364	0.249	0.309	0.284	0.362	0.399	0.445	0.269	0.328	0.280	0.339	0.281	0.340
	336	0.287	0.332	0.311	0.348	0.307	0.342	0.305	0.343	0.597	0.542	0.377	0.422	0.321	0.351	0.369	0.427	0.637	0.591	0.325	0.366	0.334	0.361	0.339	0.372
	720	0.371	0.385	0.412	0.407	0.407	0.398	0.402	0.400	1.730	1.042	0.558	0.524	0.408	0.403	0.554	0.522	0.960	0.735	0.421	0.415	0.417	0.413	0.433	0.432

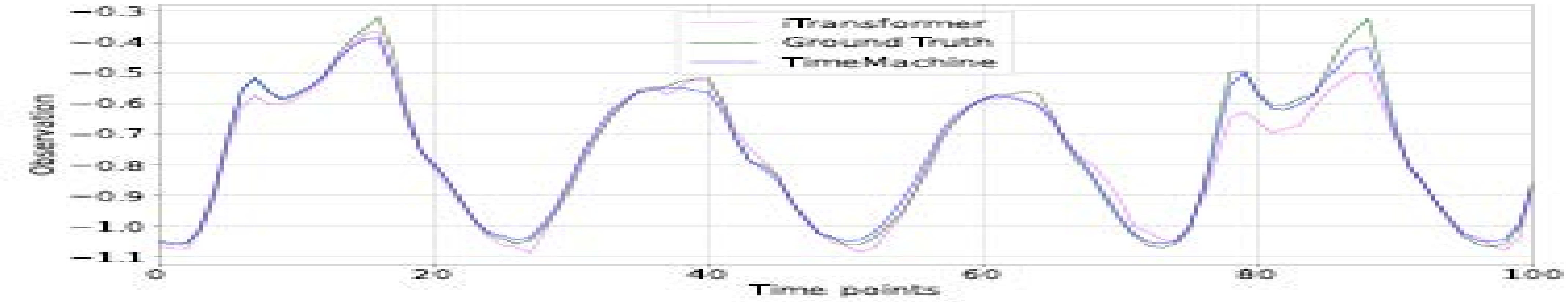
- Superior performance in almost all cases
- Effectiveness in handling varying number of channels and look-back windows

RESULTS - QUALITATIVE

Visually comparing TimeMachine with best-performing baselines



(a) Electricity

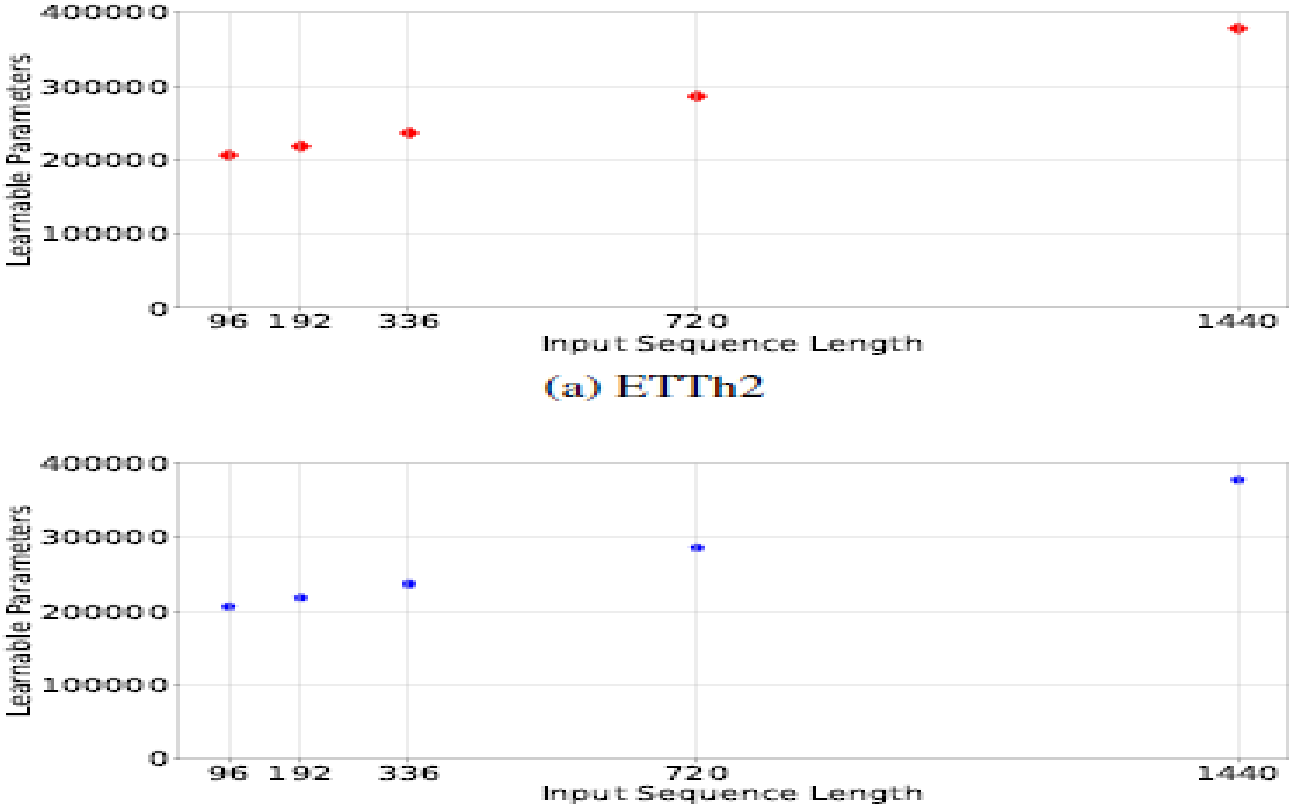


(b) Traffic

- TimeMachine aligns well with ground truth

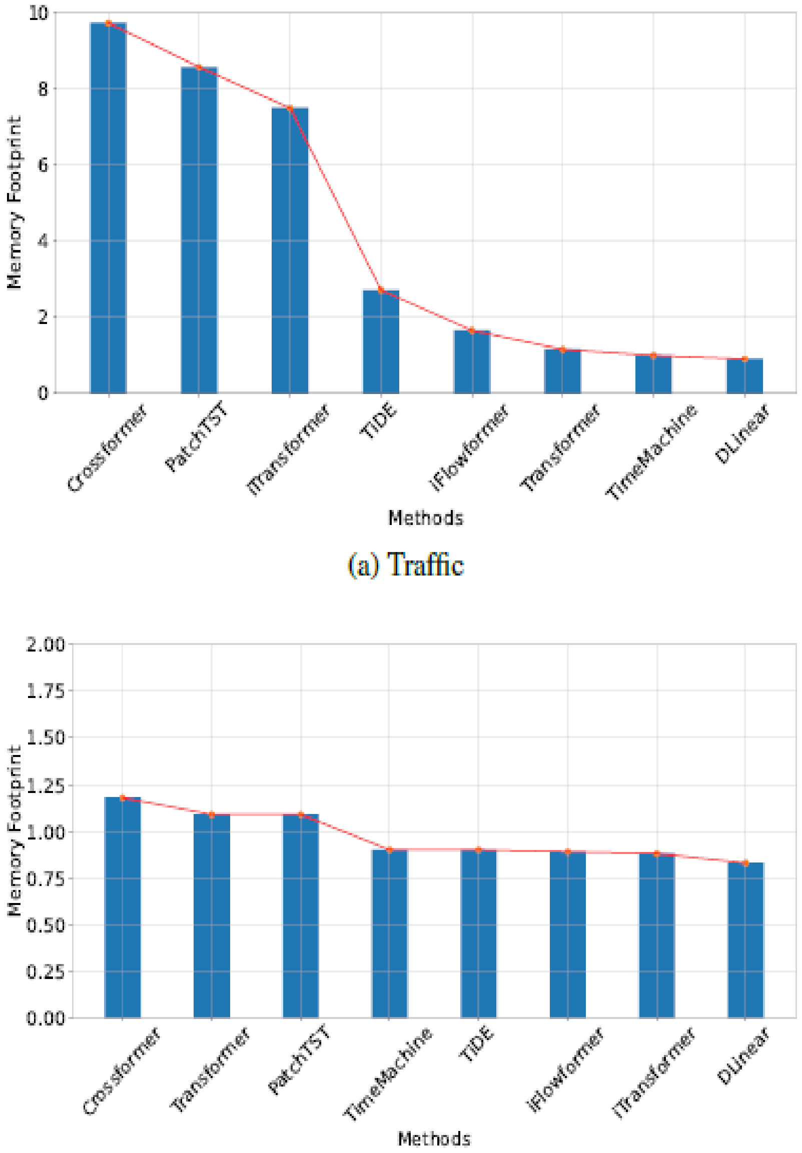
ANALYSIS - MEMORY FOOTPRINT AND SCALABILITY

- Comparison of memory footprints with baselines (Traffic: 862 channels; Weather: 21 channels)
- Linear scalability in terms of learnable parameters



(a) ETTh2

(b) Weather



(a) Traffic

(b) Weather

BRIEF CONCLUSION OF TOPIC 2

- TimeMachine: novel deep learning model for LTSF
- Superior performance, linear scalability, small memory footprints

- Potential for future exploration in self-supervised learning setting

TOPIC 3 - CROSS-VIEW DIVERSITY EMBEDDED CONSENSUS LEARNING FOR MULTI-VIEW CLUSTERING

- Importance of multi-view clustering (MVC)
  - Clustering is fundamental;
  - Data from multiple sources often need multi-view clustering
- Challenges in MVC
  - How to find consensus information from multiple views
  - Diversity not properly embedded – redundant information across views often emphasized
  - How to simultaneously learn consensus and diversity information
- Limitations of existing methods
  - Existing methods follow 2 steps: affinity matrix; spectral clustering
    - for affinity matrix, low-rank tensor recovery: tensor nuclear norm
  - Diversity is not incorporated
  - High-order neighbor information is rarely considered

RELATED WORK

Two main categories of MVC methods

- Spectral clustering (SPC)-based subspace clustering: Markov random walk
  - RMSC: transition prob. matrix for each view, common low-rank stochastic matrix by combining w. Markov mixture
  - MVC: low-rank tensor recovery from tensor nuclear norm
  - ETLMSC: essential low-rank tensor
- Graph-based subspace clustering: affinity matrix from similarity matrix
  - Focus on graph-based methods using similarity matrix

ETLMSC:

- Multi-view data: $\{\mathbf{X}^{(v)}\}_{v=1}^{V}$ where $\mathbf{X}^{(v)} \in \mathbb{R}^{d_v \times n}$
- Trans. Prob. Matrix-based tensor: $\mathcal{P} \in \mathbb{R}^{n \times n \times V}$

$$\min_{\mathcal{Z},\mathcal{E}\in\mathbb{R}^{n\times n\times V}} \|\mathcal{Z}\|_{\circledast} + \lambda\|\mathcal{E}\|_{2,1} \quad s.t. \quad \mathcal{P} = \mathcal{Z} + \mathcal{E},$$

KEY INNOVATIONS

- Recovering low-rank essential tensor from cross-order neighbor graph tensor
- Embedding auto-adjusted weighting vector for cross-view diversity and consensus
- Efficient optimization algorithm with convergence guarantee
- Superior performance over baselines

NEW METHOD: CCL-MVC

- Constructing cross-order neighbor tensor
  - Higher-order neighbor relationships
  - Fine-grained probability tensor with local structure preservation

$$\mathcal{P}_K^{(v)} = \sum_{k=1}^{K} \frac{(\mathcal{P}^{(v)})^k + ((\mathcal{P}^{(v)})^k)^T}{2}$$

- Recovering low-rank essential tensor: log-based rank and sparsity approx.

$$\min_{\mathcal{Z},\mathcal{E}} \|\mathcal{Z}\|_{®} + \lambda \|\mathcal{E}\|_{Ⓢ} \quad s.t. \quad \mathcal{P}_K = \mathcal{Z} + \mathcal{E}$$

$$\|\mathcal{Z}\|_{®} = \sum_{v=1}^{V} \sum_{i=1}^{n} \log(1 + \boldsymbol{\sigma}_i(\hat{\mathcal{Z}}^{(v)}))$$

$$\|\mathcal{E}\|_{Ⓢ} = \sum_{j=1}^{n} \log(1 + (\sum_{i=1}^{n} \sum_{v=1}^{V} \mathcal{E}_{ijv}^2)^{1/2})$$

- C Peng, Y Liu, K Kang, Y Chen, X Wu, A Cheng, Z Kang, C Chen, and Q Cheng. Hyperspectral image denoising using nonconvex local low-rank and sparse separation with spatial-spectral total variation regularization. IEEE Transactions on Geoscience and Remote Sensing, 60:1–17, 2022.
- C Peng, Z Kang, H Li, and Q Cheng. Subspace clustering using log-determinant rank approximation. ACM KDD, pp. 925–934, 2015

University of Kentucky

NEW METHOD: CCL-MVC

- Constructing consensus representation matrix

$$\min_{\mathcal{Z},\mathcal{E},w} \|\mathcal{Z}\|_{\circledR} + \lambda\|\mathcal{E}\|_{\circledS} + \alpha \sum_v \|\mathrm{Z}_0 - w_v \mathcal{Z}^{(v)}\|_F^2$$

$$s.t.\ \mathcal{P}_K = \mathcal{Z} + \mathcal{E}, w_v \geq 0, \sum_v w_v = 1,$$

Z_0: consensus affinity matrix fusing cross-view neighbor graphs

w_v: auto adjusted weights for the v-th view

- Twin learning for similarity and clustering: A unified kernel approach. Z Kang, C Peng, Q Cheng, AAAI, 2017.
- Kernel-driven similarity learning. Z Kang, C Peng, Q Cheng, Neurocomputing. 2017; 210-219.

CCL-MVC FORMULATION

- Incorporating Diversity Representation Matrix Learning
  - Auto-adjusted weighting vector for cross-view diversity
  - Embedding fusion into the model

$$\min_{\mathcal{P}_K = \mathcal{Z}+\mathcal{E}, w \geq 0, \sum_i w_i = 1} \|\mathcal{Z}\|_{\circledR} + \lambda \|\mathcal{E}\|_{\circledS}$$
$$+\alpha \sum_{v=1}^{V} \|Z_0 - w_v \mathcal{Z}^{(v)}\|_F^2 + \beta \sum_{i,j=1}^{V} w_i w_j \mathrm{Tr}((\mathcal{Z}^{(i)})^T \mathcal{Z}^{(j)}),$$

Cross-View Diversity Embedded Consensus Learning for Multi-View Clustering.
C Peng, K Zhang, Y Chen, C Chen, Q Cheng. IJCAI 2024.

OPTIMIZATION ALGORITHM

Alternating optimization with ALM

Sub-problems and solutions for each variable

- Optimization of Z:

$$\min_{\mathcal{Z}} \alpha \sum_{v} \|\mathrm{Z}_0 - w_v \mathcal{Z}^{(v)}\|_F^2 + \beta \sum_{i,j} w_i w_j \mathrm{Tr}((\mathcal{Z}^{(i)})^T \mathcal{Z}^{(j)}) + \frac{\rho}{2}\|\mathcal{Q} - \mathcal{Z} + \mathcal{Y}_1/\rho\|_F^2 + \frac{\rho}{2}\|\mathcal{P}_K - \mathcal{Z} - \mathcal{E} + \mathcal{Y}_2/\rho\|_F^2.$$

- Optimization of Q:

$$\min_{\mathcal{Q}} \|\mathcal{Q}\|_{\circledR} + \frac{\rho}{2}\|\mathcal{Q} - \mathcal{Z} + \mathcal{Y}_1/\rho\|_F^2$$

- Optimization of E:

$$\min_{\mathcal{E}} \lambda\|\mathcal{E}\|_{\circledS} + \frac{\rho}{2}\|\mathcal{P}_K - \mathcal{Z} - \mathcal{E} + \mathcal{Y}_2/\rho\|_F^2$$

- Optimization of Z_0, w: straightforward $\alpha \sum_v \|\mathrm{Z}_0 - w_v \mathcal{Z}^{(v)}\|_F^2$
- Updating of Y_1, Y_2, and ρ: standard steps

CONVERGENCE ANALYSIS

Theorem 1: Boundedness of variable sequences

Theorem 1. Let t in the superscript denote the iteration number. Under assumptions that $\sum \frac{1}{\rho^t} < \infty$ and $\sum \frac{\rho^{t+1}}{(\rho^t)^2} < \infty$, and given a bounded initialization of the variables, the variable sequences $\{\mathcal{Z}^t\}$, $\{\mathcal{E}^t\}$, $\{\mathcal{Q}^t\}$, $\{\mathrm{Z}_0^t\}$, $\{w^t\}$, $\{\mathcal{Y}_1^t\}$, and $\{\mathcal{Y}_2^t\}$ generated by our optimization algorithm are bounded.

Theorem 2: Convergence to a stationary point

Theorem 2. Let $\{\mathcal{Z}^t,\mathcal{E}^t,\mathcal{Q}^t,\mathrm{Z}_0^t,w^t,\mathcal{Y}_1^t,\mathcal{Y}_2^t\}$ be a sequence generated by our algorithm. Under assumptions that $\sum \frac{1}{\rho^t} < \infty$, $\sum \frac{\rho^{t+1}}{(\rho^t)^2} < \infty$, $\rho^t(\mathcal{Q}^{t+1} - \mathcal{Q}^t) \to 0$, and $\rho^t(\mathcal{E}^{t+1} - \mathcal{E}^t) \to 0$, the sequence $\{\mathcal{Z}^t, \mathcal{E}^t, \mathcal{Q}^t, \mathrm{Z}_0^t, w^t, \mathcal{Y}_1^t, \mathcal{Y}_2^t\}$ has at least one accumulation point. For any accumulation point, denoted as $\{\mathcal{Z}^*, \mathcal{E}^*, \mathcal{Q}^*, \mathrm{Z}_0^*, w^*, \mathcal{Y}_1^*, \mathcal{Y}_2^*\}$, $\{\mathcal{Z}^*, \mathcal{E}^*, \mathcal{Q}^*, \mathrm{Z}_0^*, w^*\}$ is a stationary point of the optimization problem in Eq. (5).

University of Kentucky

EXPERIMENTAL SETUP

- Datasets: 6 standard benchmark datasets

  - BBC-4view, BBC-Sport, Flowers, UCI-3view, StillDB, MITindoor,

- Evaluation metrics: 4 metrics

  - clustering accuracy (ACC), normalized mutual information (NMI), adjusted rand index (AR), F-Score

- Parameter settings
- Comparing our method with 17 SOTA baselines

  - CCL-MVC: Cross-view diversity embedded Consensus Learning for Multi-View Clustering

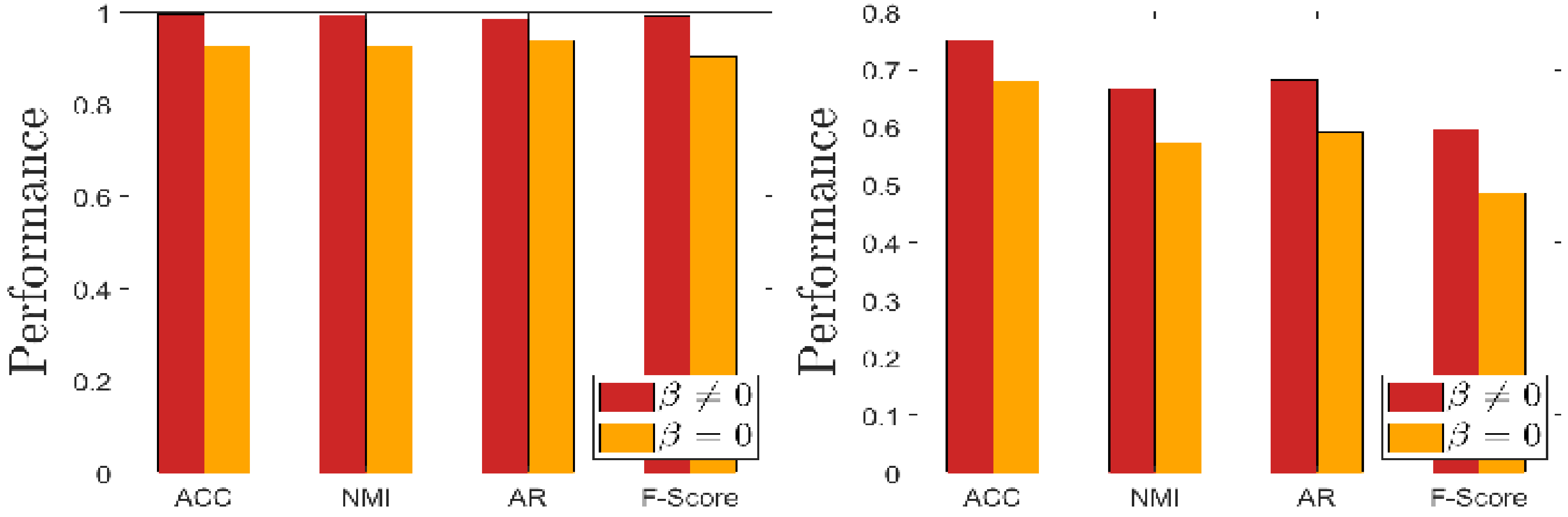
  - Superiority and stability

Datasets	BBC-4view				BBC-Sport				Flowers			
Methods	ACC	NMI	AR	F-Score	ACC	NMI	AR	F-Score	ACC	NMI	AR	F-Score
AWP	0.904±0.000	0.760±0.000	0.797±0.000	0.845±0.000	0.809±0.000	0.723±0.000	0.726±0.000	0.796±0.000	0.435±0.000	0.430±0.000	0.246±0.000	0.292±0.000
MvDSCN	0.495±0.019	0.247±0.022	0.224±0.023	0.437±0.015	0.931±0.001	0.935±0.000	0.909±0.001	0.860±0.000	0.276±0.012	0.285±0.012	0.108±0.010	0.182±0.008
OMVFC-LICAG	0.718±0.000	0.586±0.000	0.584±0.000	0.688±0.000	0.724±0.000	0.571±0.000	0.397±0.000	0.627±0.000	0.397±0.000	0.404±0.000	0.208±0.000	0.259±0.000
MLAN	0.853±0.007	0.698±0.010	0.716±0.005	0.783±0.004	0.721±0.000	0.779±0.000	0.591±0.000	0.714±0.000	0.501±0.008	0.532±0.003	0.331±0.010	0.373±0.009
GMC	0.693±0.000	0.563±0.000	0.479±0.000	0.633±0.000	0.807±0.000	0.760±0.000	0.722±0.000	0.794±0.000	0.177±0.000	0.247±0.000	0.020±0.000	0.125±0.000
UOMvSC	0.391±0.000	0.170±0.000	0.105±0.000	0.267±0.000	0.529±0.000	0.302±0.000	0.209±0.000	0.415±0.000	0.507±0.000	0.508±0.000	0.313±0.000	0.355±0.000
EOMSC-CA	0.425±0.000	0.090±0.000	0.090±0.000	0.365±0.000	0.384±0.000	0.083±0.000	0.058±0.000	0.375±0.000	0.374±0.000	0.418±0.000	0.221±0.000	0.281±0.000
LMSC	0.883±0.000	0.699±0.000	0.746±0.000	0.806±0.000	0.847±0.003	0.739±0.001	0.749±0.001	0.810±0.001	0.442±0.009	0.444±0.009	0.275±0.007	0.318±0.012
MCLES	0.819±0.000	0.637±0.000	0.662±0.000	0.742±0.000	0.921±0.000	0.802±0.000	0.795±0.000	0.945±0.000	0.469±0.000	0.516±0.000	0.337±0.000	0.390±0.000
FPMVS-CAG	0.323±0.000	0.030±0.000	0.037±0.000	0.276±0.000	0.406±0.000	0.106±0.000	0.103±0.000	0.304±0.000	0.272±0.000	0.356±0.000	0.182±0.000	0.248±0.000
CLR-MVP	0.626±0.000	0.476±0.000	0.437±0.000	0.597±0.000	0.925±0.000	0.815±0.000	0.851±0.000	0.886±0.000	0.520±0.000	0.533±0.001	0.365±0.001	0.403±0.001
t-SVD-MSC	0.858±0.001	0.685±0.002	0.725±0.002	0.789±0.001	0.879±0.000	0.765±0.000	0.784±0.000	0.834±0.000	**0.836±0.005**	0.852±0.002	**0.766±0.002**	**0.780±0.002**
SM^2SC	0.934±0.008	0.812±0.001	0.853±0.003	0.887±0.006	**0.982±0.000**	0.937±0.000	**0.952±0.000**	**0.963±0.000**	0.442±0.008	0.453±0.005	0.276±0.007	0.319±0.007
ETLMSC	0.872±0.094	0.826±0.028	0.811±0.082	0.855±0.063	0.959±0.086	**0.972±0.058**	0.949±0.107	0.961±0.081	0.811±0.066	**0.874±0.025**	0.763±0.057	0.778±0.054
LMVSC	0.480±0.000	0.242±0.000	0.403±0.000	0.380±0.000	0.517±0.000	0.382±0.000	0.151±0.000	0.394±0.000	0.360±0.000	0.385±0.000	0.198±0.000	0.246±0.000
E^2OMVC	0.849±0.000	0.707±0.000	0.713±0.000	0.783±0.000	0.971±0.000	0.903±0.000	0.920±0.000	0.940±0.000	0.490±0.000	0.477±0.000	0.256±0.000	0.306±0.000
RMSL	**0.943±0.009**	**0.831±0.005**	**0.862±0.004**	**0.894±0.002**	0.972±0.002	0.905±0.005	0.931±0.002	0.947±0.004	0.511±0.006	0.490±0.007	0.332±0.010	0.372±0.005
CCL-MVC	**0.984±0.000**	**0.951±0.000**	**0.962±0.000**	**0.971±0.000**	**0.996±0.000**	**0.986±0.000**	**0.991±0.000**	**0.993±0.000**	**0.890±0.053**	**0.892±0.019**	**0.832±0.052**	**0.852±0.048**

Datasets	UCI-3view				StillDB				MITindoor			
Methods	ACC	NMI	AR	F-Score	ACC	NMI	AR	F-Score	ACC	NMI	AR	F-Score
AWP	0.806±0.000	0.842±0.000	0.759±0.000	0.785±0.000	0.306±0.000	0.093±0.000	0.058±0.000	0.223±0.000	0.499±0.000	0.629±0.000	0.317±0.000	0.329±0.000
MvDSCN	0.308±0.011	0.299±0.013	0.158±0.009	0.281±0.006	0.377±0.023	0.245±0.020	0.169±0.003	0.320±0.015	0.084±0.003	0.182±0.004	0.014±0.002	0.037±0.001
OMVFC-LICAG	0.833±0.000	0.811±0.000	0.731±0.000	0.759±0.000	0.376±0.000	0.129±0.000	0.087±0.000	0.273±0.000	0.319±0.000	0.453±0.000	0.157±0.000	0.171±0.000
MLAN	0.874±0.000	0.910±0.000	0.847±0.000	0.864±0.000	0.349±0.000	0.138±0.000	0.098±0.000	0.272±0.000	0.232±0.010	0.408±0.012	0.012±0.009	0.041±0.003
GMC	0.736±0.000	0.815±0.000	0.678±0.000	0.713±0.000	0.251±0.000	0.078±0.000	0.005±0.000	0.278±0.000	0.099±0.000	0.204±0.000	0.003±0.000	0.032±0.000
UOMvSC	**0.981±0.000**	0.956±0.000	**0.958±0.000**	**0.962±0.000**	0.328±0.000	0.131±0.000	0.084±0.000	0.246±0.000	0.344±0.000	0.506±0.000	0.064±0.000	0.088±0.000
EOMSC-CA	0.545±0.000	0.673±0.000	0.459±0.000	0.533±0.000	0.308±0.000	0.127±0.000	0.085±0.000	0.245±0.000	0.147±0.000	0.298±0.000	0.042±0.000	0.066±0.000
LMSC	0.893±0.000	0.815±0.000	0.783±0.000	0.805±0.000	0.327±0.003	0.136±0.003	0.084±0.011	0.269±0.005	0.384±0.006	0.506±0.005	0.243±0.005	0.254±0.004
MCLES	0.941±0.004	0.891±0.008	0.877±0.009	0.889±0.008	0.338±0.000	0.153±0.000	0.098±0.000	0.264±0.000	—	—	—	—
FPMVS-CAG	0.722±0.000	0.744±0.000	0.645±0.000	0.683±0.000	0.328±0.000	0.124±0.000	0.089±0.000	0.251±0.000	0.204±0.000	0.390±0.000	0.085±0.000	0.108±0.000
CLR-MVP	0.965±0.000	0.920±0.001	0.924±0.001	0.932±0.001	0.337±0.002	0.127±0.003	0.095±0.001	0.273±0.002	—	—	—	—
t-SVD-MSC	0.830±0.000	0.884±0.005	0.786±0.003	0.800±0.004	0.347±0.010	0.130±0.004	0.088±0.003	0.255±0.004	0.684±0.005	0.750±0.007	0.555±0.005	0.562±0.008
SM^2SC	0.961±0.001	0.914±0.001	0.914±0.001	0.923±0.001	0.452±0.002	0.336±0.002	0.312±0.002	0.370±0.002	0.477±0.009	0.583±0.006	0.321±0.006	0.332±0.006
ETLMSC	0.958±0.078	**0.977±0.028**	0.953±0.069	0.958±0.062	**0.604±0.043**	**0.520±0.015**	**0.423±0.029**	**0.523±0.024**	**0.775±0.040**	**0.899±0.011**	**0.729±0.037**	**0.733±0.036**
LMVSC	0.790±0.000	0.756±0.000	0.643±0.000	0.681±0.000	0.317±0.000	0.189±0.000	0.066±0.000	0.239±0.000	0.371±0.000	0.522±0.000	0.112±0.000	0.132±0.000
E^2OMVC	0.974±0.000	0.941±0.000	0.943±0.000	0.949±0.000	0.321±0.000	0.130±0.000	0.089±0.000	0.265±0.000	0.404±0.000	0.550±0.000	0.173±0.000	0.191±0.000
RMSL	0.578±0.013	0.511±0.014	0.407±0.017	0.474±0.007	0.356±0.003	0.131±0.001	0.090±0.002	0.243±0.001	0.279±0.004	0.372±0.003	0.125±0.005	0.139±0.002
CCL-MVC	**0.988±0.037**	**0.996±0.013**	**0.988±0.037**	**0.990±0.033**	**0.752±0.077**	**0.682±0.042**	**0.597±0.079**	**0.666±0.066**	**0.860±0.038**	**0.946±0.011**	**0.842±0.038**	**0.844±0.037**

ABLATION STUDY & CONVERGENCE

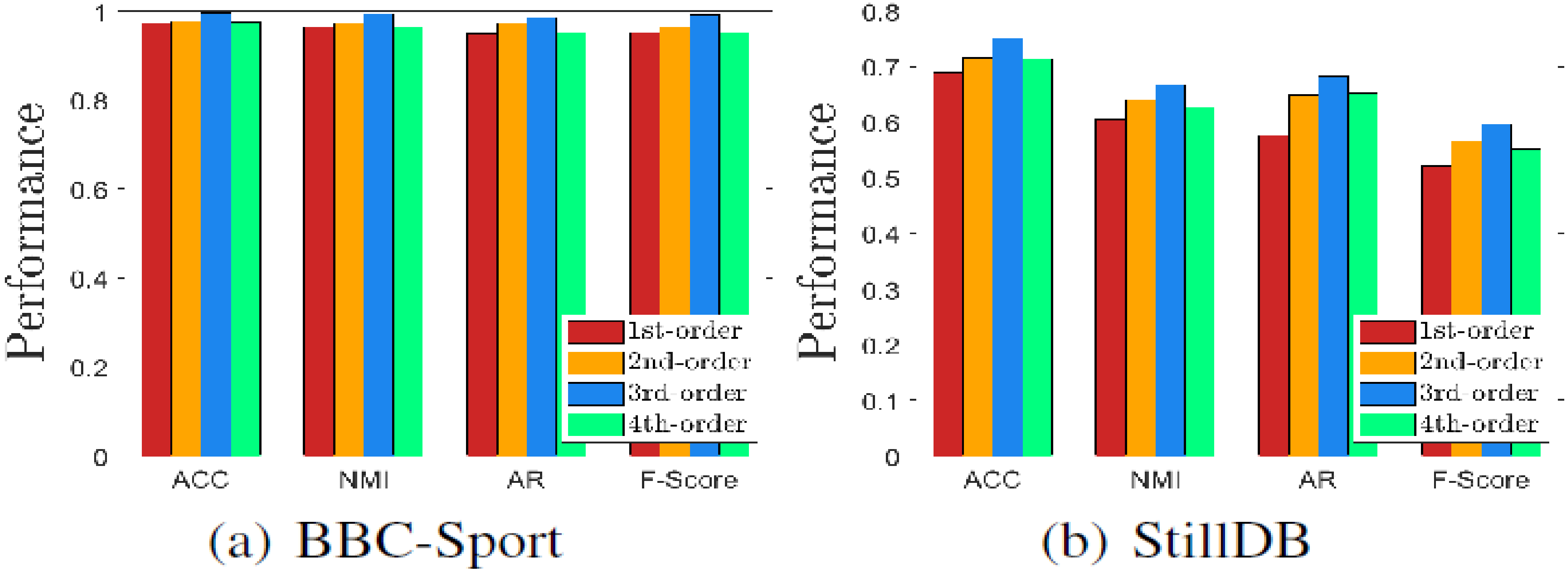
Significance of cross-view diversity



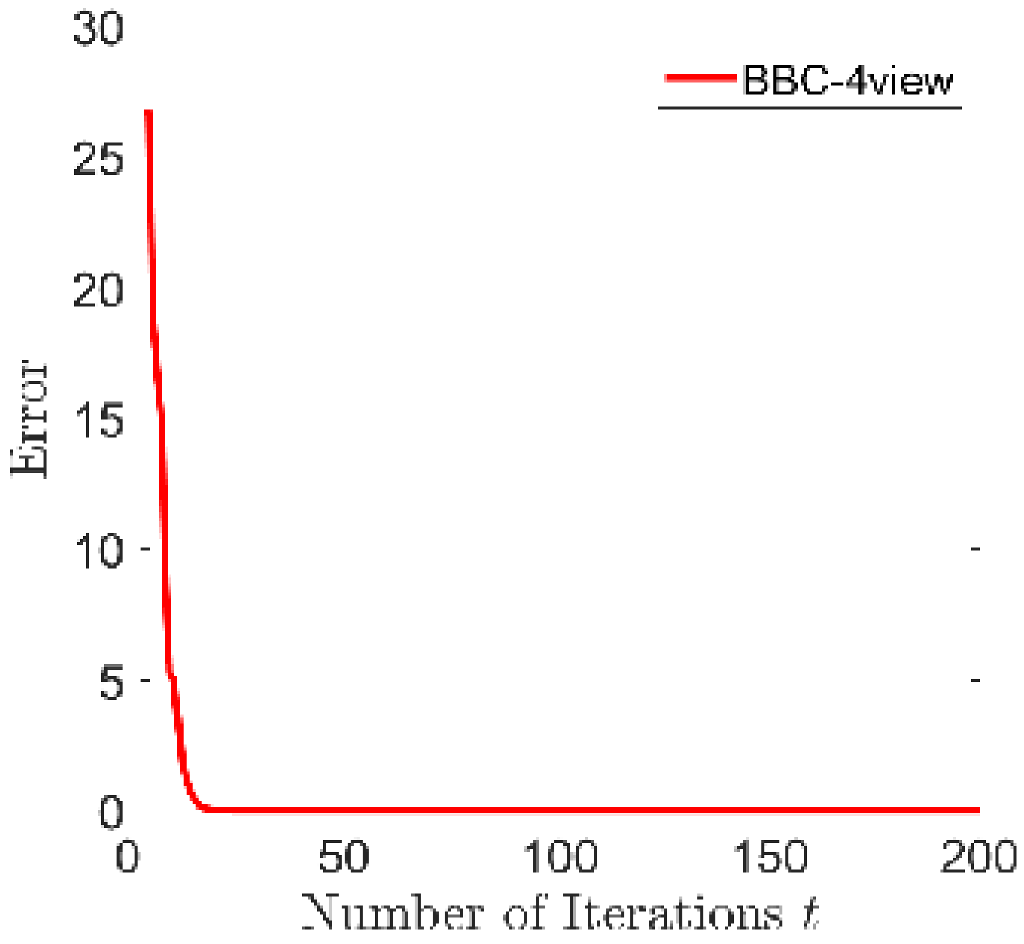
(a) BBC-Sport (b) StillDB

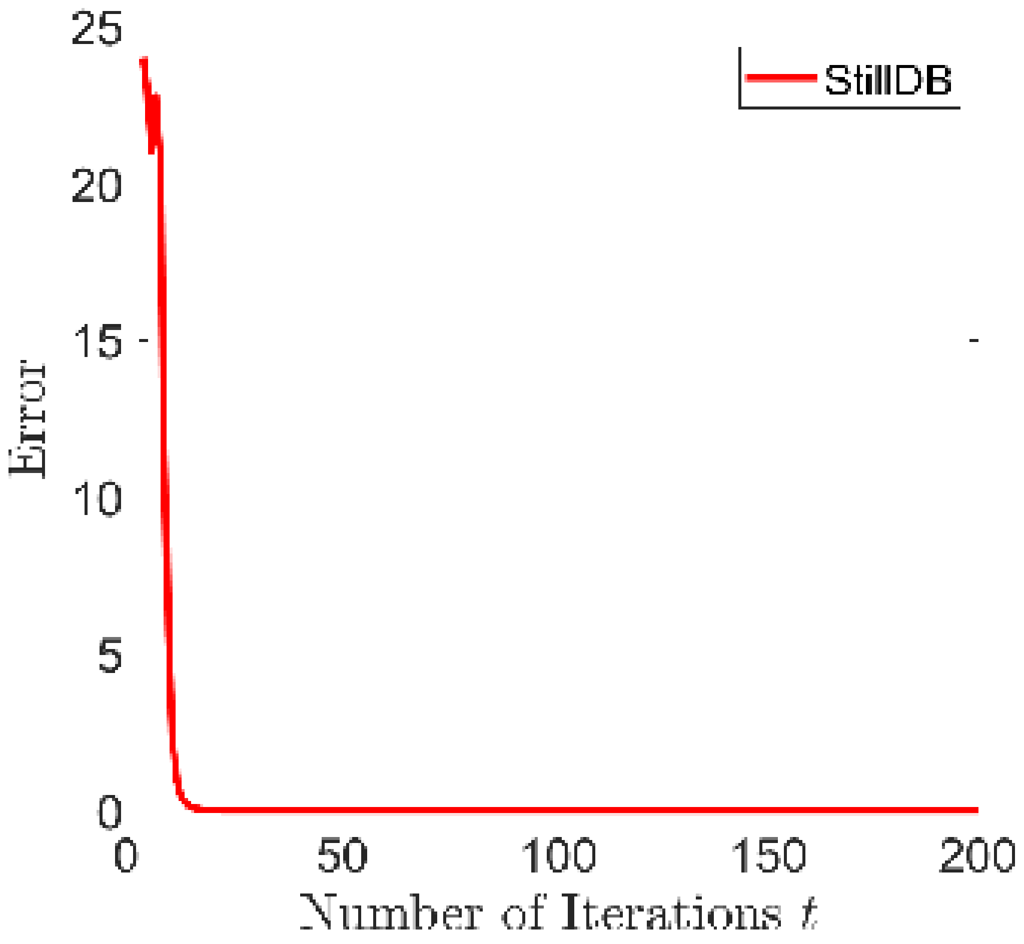
University of Kentucky

CROSS-ORDER NEIGHBOR INFORMATION



CONVERGENCE CURVES ON SAMPLE DATASETS





BRIEF CONCLUSION OF TOPIC 3

- CCL-MVC for enhanced MVC with diversity and cross-order info
- Efficient optimization with convergence guarantee
- Superior performance experimentally validated

Future Work:

- More accurate approximation of tensor rank

TOPIC 4 – NEW DEEP LEARNING MODEL FOR TIME SERIES CLASSIFICATION

Motivation

- Importance of time series classification (TSC)
  - Structured format; mixed types; predominant in various domains
- Challenges with existing deep learning models for TSC
  - Classical methods (DTW, XGBoost): accuracy, long-range dependency
  - Deep learning models (Rocket, LSTM, TCN): accuracy, LRD
  - Transformer-based models (Informer, AutoFormer, FlowFormer, FedFormer): efficiency, scalability
- Need for an efficient, accurate, scalable approach
  - Also, existing models does not consider inversion invariance

NEW MODEL: TSCMAMBA

Key Innovations

- Captures features that are robust to time shifting and time inversion

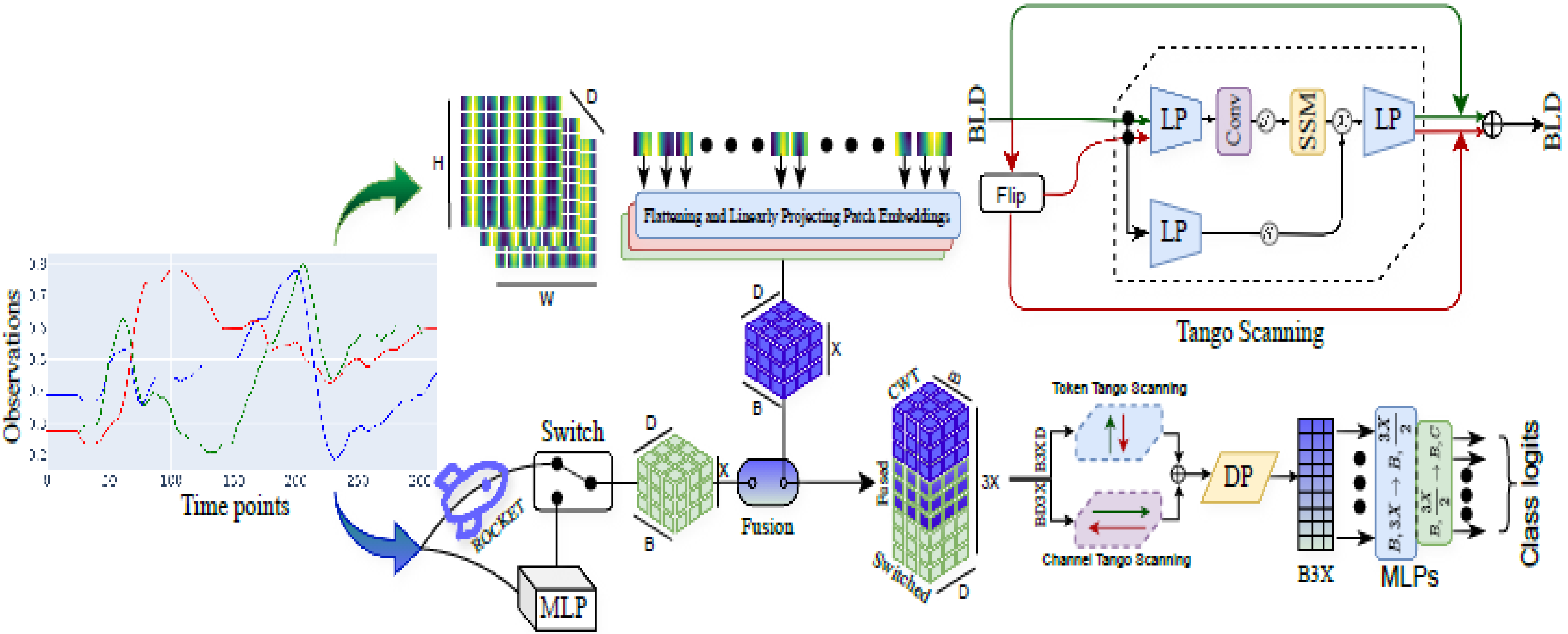
- Use multi-view framework to integrate time-frequency features at global and local scales for classification

- First to leverage SSM modules (Mamba) for TSC with linear scalability and small memory footprints

- Creates a new Mamba scanning scheme (Tango scanning): capturing inversion-invariant features

TSCMamba: Mamba Meets Multi-View Learning for Time Series Classification, MA Atik, Q Cheng. 2024. arXiv preprint arXiv:2406.04419

MODEL ARCHITECTURE



University of Kentucky

DATASETS FOR EVALUATIONS

	Datasets	Channels	Length	Train	Test	Classes	Domain
Benchmark datasets	EthanolConcentration (EC)	3	1751	261	263	4	Alcohol Industry
	FaceDetection (FD)	144	62	5890	3524	2	Face (250Hz)
	Handwriting (HW)	3	152	150	850	26	Smart Watch
	Heartbeat (HB)	61	405	204	205	2	Clinical
	JapaneseVowels (JV)	12	29	270	370	9	Audio
	PEMS-SF (PS)	963	144	267	173	7	Transportation
	SelfRegulationSCP1 (SCP1)	6	896	268	293	2	Health (256Hz)
	SelfRegulationSCP2 (SCP2)	7	1152	200	180	2	Health (256Hz)
	SpokenArabicDigits (SA)	13	93	6599	2199	10	Voice (11025Hz)
	UWaveGestureLibrary (UG)	3	315	120	320	8	Gesture
Additional datasets	AtrialFibrillation (AF)	2	640	15	15	3	ECG
	BasicMotions (BM)	6	100	40	40	4	Human Activity Recognition
	Cricket (CR)	6	1197	108	72	12	Human Activity Recognition
	FingerMovements (FM)	28	50	316	100	2	EEG
	HandMovementDirection (HMD)	10	400	160	74	4	EEG
	MotorImagery (MI)	64	3000	278	100	2	EEG
	PenDigits (PD)	2	8	7494	3498	10	Motion
	PhonemeSpectra (PHS)	11	217	3315	3353	39	Audio
	RacketSports (RS)	6	30	151	152	4	Human Activity Recognition
	StandWalkJump (SWJ)	4	2500	12	15	3	ECG

EXPERIMENTAL RESULTS (20 BASELINES)

Classification Accuracy (%). The . symbol in Transformer models denotes $*$former used.
The best average result and rank: in **bold**; the second best: underlined.
The ranks: Wilcoxon signed-rank test (lower is better).

Datasets	DTW (1994)	XGBoost (2016)	Rocket (2020)	LSTM (1997)	LSTNet (2018)	LSSL (2022)	TCN (2019)	Trans (2017)	Re. (2020)	In. (2021)	Pyra (2021)	Auto. (2021)	Station. (2022)	FED. (2022)	ETS. (2022)	Flow. (2022)	DLinear (2023)	LightTS. (2022a)	TimesNet (2023)	TSLANet (2024)	TSCMamba Ours
EC	32.3	43.7	45.2	32.3	39.9	31.1	28.9	32.7	31.9	31.6	30.8	31.6	32.7	31.2	28.1	33.8	32.6	29.7	35.7	30.4	62.0
FD	52.9	63.3	64.7	57.7	65.7	66.7	52.8	67.3	68.6	67.0	65.7	68.4	68.0	66.0	66.3	67.6	68.0	67.5	68.6	66.8	69.4
HW	28.6	15.8	58.8	15.2	25.8	24.6	53.3	32.0	27.4	32.8	29.4	36.7	31.6	28.0	32.5	33.8	27.0	26.1	32.1	57.9	53.3
HB	71.7	73.2	75.6	72.2	77.1	72.7	75.6	76.1	77.1	80.5	75.6	74.6	73.7	73.7	71.2	77.6	75.1	75.1	78.0	77.6	76.6
JV	94.9	86.5	96.2	79.7	98.1	98.4	98.9	98.7	97.8	98.9	98.4	96.2	99.2	98.4	95.9	98.9	96.2	96.2	98.4	99.2	97.0
PS	71.1	98.3	75.1	39.9	86.7	86.1	68.8	82.1	82.7	81.5	83.2	82.7	87.3	80.9	86.0	83.8	75.1	88.4	89.6	83.8	90.2
SCP1	77.7	84.6	90.8	68.9	84.0	90.8	84.6	92.2	90.4	90.1	88.1	84.0	89.4	88.7	89.6	92.5	87.3	89.8	91.8	91.8	92.5
SCP2	53.9	48.9	53.3	46.6	52.8	52.2	55.6	53.9	56.7	53.3	53.3	50.6	57.2	54.4	55.0	56.1	50.5	51.1	57.2	61.7	66.7
SA	96.3	69.6	71.2	31.9	100.0	100.0	95.6	98.4	97.0	100.0	99.6	100.0	100.0	100.0	100.0	98.8	81.4	100.0	99.0	99.9	99.0
UG	90.3	75.9	94.4	41.2	87.8	85.9	88.4	85.6	85.6	85.6	83.4	85.9	87.5	85.3	85.0	86.6	82.1	80.3	85.3	91.3	93.8
Avg.	67.0	66.0	72.5	48.6	71.8	70.9	70.3	71.9	71.5	72.1	70.8	71.1	72.7	70.7	71.0	73.0	67.5	70.4	73.6	<u>76.04</u>	**80.05**
Rank	15.20	15.55	10.25	19.55	10.40	11.70	12.40	9.40	9.95	8.90	12.80	11.50	7.30	12.40	12.85	6.45	14.60	12.65	<u>6.40</u>	<u>6.40</u>	**4.35**

COMPUTATIONAL EFFICIENCY

FLOPs comparison among the top performing methods.
The values: in GigaFLOPS (G) or TeraFlops (T), 1 TFLOPs=1000 GFLOPs
A lower value: better computational efficiency

Methods	EC	FD	HW	HB	JV	PS	SCP1	SCP2	SA	UG
Flow. Wu et al. (2022)	1.06T	37.97G	92.21G	246.37G	15.76G	94.02G	542.64G	697.74G	50.33G	190.82G
TimesNet. Wu et al. (2023)	1.11T	161.93G	115.88G	182.69G	48.15G	**74.18G**	503.62G	2.33T	26.00G	247.73G
TSCMamba (Ours)	**1.69G**	**11.53G**	**27.24G**	**8.39G**	**12.33G**	2.84T	**3.42G**	**11.11G**	**0.78G**	**13.86G**

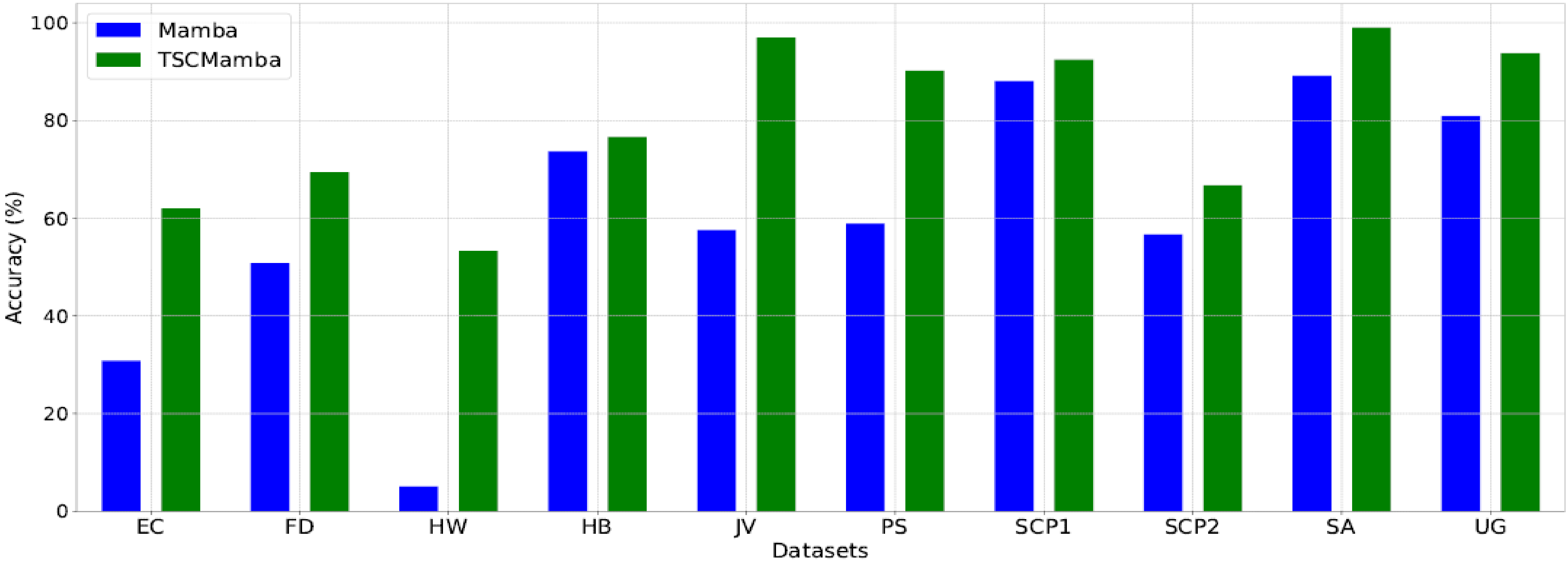
University of Kentucky

ADDITIONAL EXPERIMENTAL RESULTS

Additional classification results on the UEA datasets in accuracy (as %).
The ranks: Wilcoxon signed-rank test (lower is better).

Dataset	TSCMamba Ours	TSLANet (2024)	GPT4TS (2023)	TimesNet (2023)	ROCKET (2020)	CrossF. (2023)	PatchTST (2023)	MLP (2023)	TS-TCC (2021)	TS2VEC (2022)
AtrialFibrillation	**67.00**	40.00	33.33	33.33	20.00	46.66	53.33	46.66	33.33	53.33
BasicMotions	**100.00**	100.00	92.50	100.00	100.00	90.00	92.50	85.00	100.00	92.50
Cricket	**98.61**	98.61	8.33	87.50	98.61	84.72	84.72	91.67	93.06	65.28
FingerMovements	**69.00**	61.00	57.00	59.38	61.00	64.00	62.00	64.00	44.00	51.00
HandMovementDirection	**71.62**	52.70	18.92	50.00	50.00	58.11	58.11	58.11	64.86	32.43
MotorImagery	**62.00**	62.00	50.00	51.04	53.00	61.00	61.00	61.00	47.00	47.00
PenDigits	98.54	98.94	97.74	98.19	97.34	93.65	**99.23**	92.94	98.51	97.40
PhonemeSpectra	24.66	17.75	3.01	18.24	17.60	7.55	11.69	7.10	**25.92**	8.23
RacketSports	**91.45**	90.79	76.97	82.64	86.18	81.58	84.21	78.95	84.87	74.34
StandWalkJump	**73.33**	46.67	33.33	53.33	46.67	53.33	60.00	60.00	40.00	46.67
Average	**75.62**	66.85	47.11	63.36	63.04	64.06	66.68	64.54	63.15	56.82
Rank	**1.65**	3.90	8.60	5.70	5.70	6.00	4.35	5.95	5.45	7.70

COMPARISON WITH MAMBA



TSCMamba (using proposed Tango scanning) in comparison with directly applying regular Mamba module.

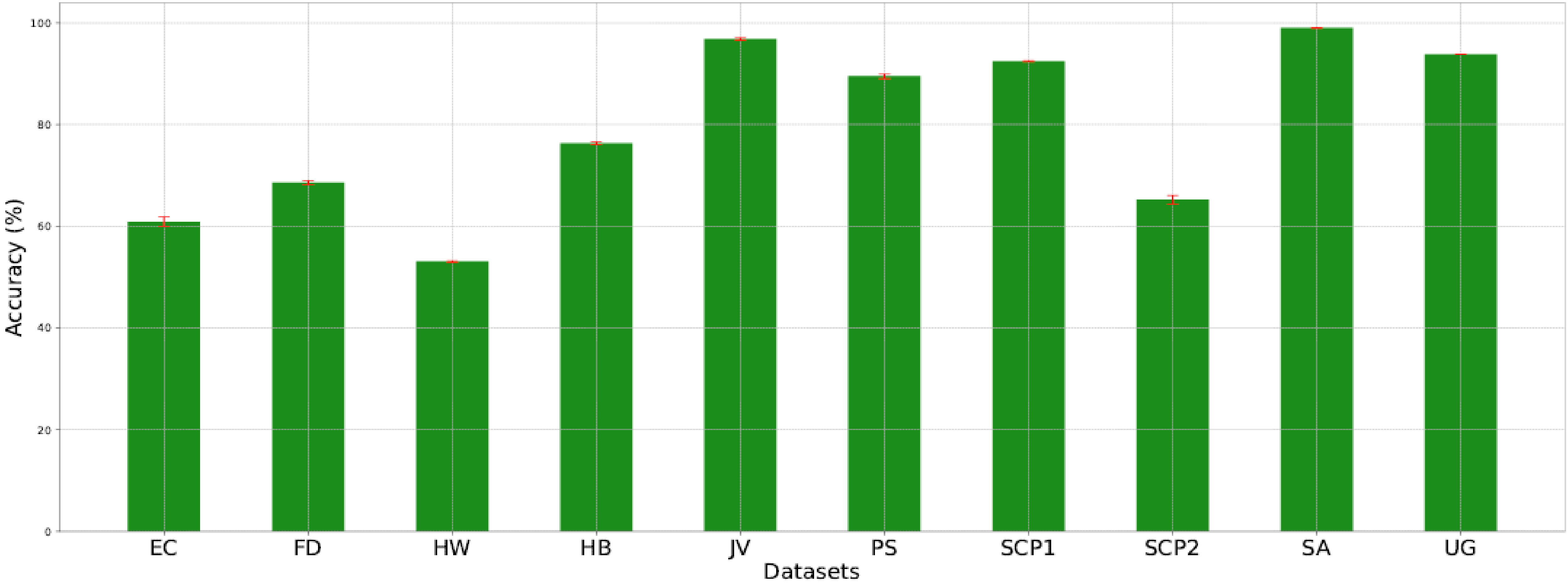
University of Kentucky

ABLATION STUDY (COMPONENT-WISE)

Ablation experiments on particular components in our method.

Mamba	Avg.Pool	ROCKET	AF	EC	FD	HW	HB	JV	PS	SCP1	SCP2	SA	UG	Avg.
✓	✓	✓	✓	62.0	57.0	53.3	74.1	93.0	90.2	92.5	66.7	94.1	93.8	77.67
✗	✓	✓	✓	33.1	63.2	34.1	73.2	85.4	81.5	86.7	57.2	74.0	89.1	67.75
✓	✗	✓	✓	31.6	64.2	52.0	74.1	94.1	63.0	86.7	60.6	96.7	92.8	71.58
✓	✓	✗	✓	31.6	69.4	24.8	76.6	97.0	87.3	91.8	58.3	97.6	86.2	72.06
✓	✓	✓	✗	30.0	51.5	49.3	72.7	91.4	84.4	88.7	58.9	90.0	90.3	70.72

MEAN AND STD DEV OVER DIFFERENT RUNS



Performance of TSCMamba over 5 random runs.
Mean performance: green bars,
Standard deviation: red error bars (very small).

BRIEF SUMMARY OF TIME SERIES CLASSIFICATION

- We create a new model for time series classification, TSCMamba
- It leverages a new scanning scheme for Mamba to take advantage of inversion invariance
- It shows better performance than SOTA models (> 10 models) on 20 benchmark time series datasets (in both averaged acc. and rank)
- It is efficient - typically uses a small fraction of FLOPS than other top performing models

REFERENCES

1. MA Ahamed, Q Cheng (2024). MambaTab: A plug-and-play model for learning tabular data. IEEE MIPR. Code available at GitHub.
2. MA Ahamed, Q Cheng (2024). TimeMachine: A Time Series is Worth 4 Mambas for Long-term Forecasting. ECAI. Code Available at GitHub.
3. C Peng, Y Liu, K Kang, Y Chen, X Wu, A Cheng, Z Kang, C Chen, and Q Cheng (2022). Hyperspectral image denoising using nonconvex local low-rank and sparse separation with spatial-spectral total variation regularization. IEEE Transactions on Geoscience and Remote Sensing, 60:1–17.
4. C Peng, Z Kang, H Li, and Q Cheng (2015). Subspace clustering using log-determinant rank approximation. ACM KDD, pp. 925–934.
5. Z Kang, C Peng, Q Cheng (2017). Twin learning for similarity and clustering: A unified kernel approach. AAAI.
6. Z Kang, C Peng, Q Cheng (2017). Kernel-driven similarity learning. Neurocomputing. 210-219.

REFERENCES

7. C Peng, K Zhang, Y Chen, C Chen, Q Cheng (2024). Cross-View Diversity Embedded Consensus Learning for Multi-View Clustering. IJCAI.

8. MA Atik, Q Cheng (2024). TSCMamba: Mamba Meets Multi-View Learning for Time Series Classification. arXiv preprint arXiv:2406.04419

Thank you.