Novelty Assessment Report

Paper: Weight-Space Linear Recurrent Neural Networks

PDF URL: https://openreview.net/pdf?id=zHdKaF3ZM7

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2025-12-27

Abstract

We introduce WARP (**W**eight-space **A**daptive **R**ecurrent **P**rediction), a simple yet powerful model that unifies weight-space learning with linear recurrence to redefine sequence modeling. Unlike conventional recurrent neural networks (RNNs) which collapse temporal dynamics into fixed-dimensional hidden states, WARP explicitly parametrizes its hidden state as the weights and biases of a distinct auxiliary neural network, and uses input differences to drive its recurrence. This brain-inspired formulation enables efficient gradient-free adaptation of the auxiliary network at test-time, in-context learning abilities, and seamless integration of domain-specific physical priors. Empirical validation shows that WARP matches or surpasses state-of-the-art baselines on diverse classification tasks, featuring in the top three in 4 out of 6 real-world challenging datasets. Furthermore, extensive experiments across sequential image completion, multivariate time series forecasting, and dynamical system reconstruction demonstrate its expressiveness and generalization capabilities. Remarkably, a physics-informed variant of our model outperforms the next best model by more than 10x. Ablation studies confirm the architectural necessity of key components, solidifying weight-space linear RNNs as a transformative paradigm for adaptive machine intelligence.

Core Task Landscape

This paper addresses: **Sequence Modeling with Weight-Space Linear Recurrence**

A total of **50 papers** were analyzed and organized into a taxonomy with **22 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Linear Recurrent Architectures and Mechanisms**
- **Training and Optimization Methods**
- **Theoretical Foundations and Analysis**
- **Domain-Specific Applications**
- **Efficiency and Deployment Optimization**
- **Representation Learning and Meta-Modeling**
- **Auxiliary Methods and Baselines**

Complete Taxonomy Tree

- Sequence Modeling with Weight-Space Linear Recurrence Survey Taxonomy
- Linear Recurrent Architectures and Mechanisms
  - Gated Linear Recurrence Models (3 papers)
  - [1] Griffin: Mixing gated linear recurrences with local attention for efficient language models (De, 2024) View paper
  - [5] GateLoop: Fully Data-Controlled Linear Recurrence for Sequence Modeling (Katsch, 2023) View paper
  - [11] Hierarchically gated recurrent neural network for sequence modeling (Qin Zhen, 2023) View paper
  - State Space Models and Structured Recurrence (4 papers)
  - [4] Mamba: Linear-Time Sequence Modeling with Selective State Spaces (Gu, 2023) View paper
  - [10] SKOLR: Structured Koopman Operator Linear RNN for Time-Series Forecasting (Zhang Yi-tian, 2025) View paper
  - [16] On Structured State-Space Duality (Hu, 2025) View paper
  - [32] Linearly recurrent autoencoder networks for learning dynamics (Otto, 2019) View paper
  - Hybrid Linear-Attention Architectures (3 papers)
  - [19] Linear-MoE: Linear Sequence Modeling Meets Mixture-of-Experts (Zhu Tong, 2025) View paper
  - [33] GoldFinch: High Performance RWKV/Transformer Hybrid with Linear Pre-Fill and Extreme KV-Cache Compression (Goldstein, 2024) View paper
  - [44] Unifying Linear-Time Attention via Latent Probabilistic Modelling (Rares Dolga, 2024) View paper
  - Weight-Space and Meta-Learning Recurrence ★ (3 papers)
  - [0] Weight-Space Linear Recurrent Neural Networks (Anon et al., 2026) View paper
  - [20] MesaNet: Sequence Modeling by Locally Optimal Test-Time Training (von Oswald, 2025) View paper
  - [24] Longhorn: State Space Models are Amortized Online Learners (Liu Bo, 2024) View paper
  - Classical Recurrent Architectures (3 papers)
  - [8] Resurrecting recurrent neural networks for long sequences (Orvieto, 2023) View paper
  - [21] Input-output HMMs for sequence processing (Yoshua Bengio, 1996) View paper
  - [28] Long short-term memory recurrent neural network architectures for large scale acoustic modeling. (HaÅim Sak, 2014) View paper
- Training and Optimization Methods
  - Curriculum and Scheduled Training (1 papers)

- [22] Scheduled sampling for sequence prediction with recurrent neural networks (Samy Bengio, 2015) View paper
- Evolutionary and Hybrid Optimization (1 papers)
- [23] Evolino: Hybrid neuroevolution/optimal linear search for sequence prediction (JÃ¼rgen Schmidhuber, 2005) View paper
- Test-Time and Online Adaptation (1 papers)
- [17] Test time learning for time series forecasting (Chen Shi-chu, 2024) View paper
- Linearization and Model Transformation (1 papers)
- [18] Liger: Linearizing Large Language Models to Gated Recurrent Structures (Hu Jiaxi, 2025) View paper

- Theoretical Foundations and Analysis
  - Expressiveness and Universality (2 papers)
  - [34] Universality of Linear Recurrences Followed by Non-linear Projections: Finite-Width Guarantees and Benefits of Complex Eigenvalues (Orvieto, 2023) View paper
  - [40] Universal neural functionals (Zhou, 2024) View paper
  - Memory and Projection Mechanisms (1 papers)
  - [14] Hippo: Recurrent memory with optimal polynomial projections (Gu, 2020) View paper
  - Unified Frameworks and Taxonomies (3 papers)
  - [6] Back to recurrent processing at the crossroad of transformers and state-space models (Matteo Tiezzi, 2025) View paper
  - [12] On the Resurgence of Recurrent Models for Long Sequences--Survey and Research Opportunities in the Transformer Era (Tiezzi, 2024) View paper
  - [36] Unlocking the Secrets of Linear Complexity Sequence Model from A Unified Perspective (Qin Zhen, 2024) View paper
- Domain-Specific Applications
  - Sequential Recommendation Systems (3 papers)
  - [2] Linear recurrent units for sequential recommendation (Zhenrui Yue, 2024) View paper
  - [3] Behavior-dependent linear recurrent units for efficient sequential recommendation (Chengkai Liu, 2024) View paper
  - [48] Temporal Linear Item-Item Model for Sequential Recommendation (Seongmin Park, 2024) View paper
  - Time Series Forecasting (4 papers)
  - [9] Bidirectional linear recurrent models for sequence-level multisource fusion (Liu, 2025) View paper
  - [27] A TCN-Linear Hybrid Model for Chaotic Time Series Forecasting (Mengjiao Wang, 2024) View paper
  - [29] Mixture-of-Linear-Experts for Long-term Time Series Forecasting (Ronghao Ni, 2023) View paper
  - [41] A Financial Time-Series Prediction Model Based on Multiplex Attention and Linear Transformer Structure (Caosen Xu, 2023) View paper
  - Language Modeling and Context Processing (1 papers)
  - [7] Resona: Improving Context Copying in Linear Recurrence Models with Retrieval (Wang, 2025) View paper
  - Vision and Spatial Sequence Modeling (1 papers)
  - [39] V"Mean"ba: Visual State Space Models only need 1 hidden dimension (Chiang, 2024) View paper
  - Scientific Computing and Dynamical Systems (3 papers)
  - [31] Regularized dynamic mode decomposition algorithm for time sequence predictions (Xiao-Yang Xie, 2024) View paper
  - [37] Caudits: Causal disentangled domain adaptation of multivariate time series (Sun, 2024) View paper
  - [38] Assessment of unsteady flow predictions using hybrid deep learning based reduced-order models (Bukka, 2021) View paper
  - Specialized Sequence Tasks (6 papers)
  - [15] Deep symbolic regression for recurrent sequences (d'Ascoli, 2022) View paper
  - [25] An integrated mediapipe-optimized GRU model for Indian sign language recognition (David Jonathan Osterhouse, 2022) View paper
  - [45] Deep Recurrent Neural Networks for Sequential Phenotype Prediction in Genomics (Pouladi, 2015) View paper
  - [47] LuKAN: A Kolmogorov-Arnold Network Framework for 3D Human Motion Prediction (Hamza, 2025) View paper
  - [49] Enhanced Multi-Vehicle Trajectory Prediction via an Extended Temporal Sequence Fusion Attention Network (Dengyu Xiao, 2025) View paper
  - [50] Linear Diffusion Networks (Fein-Ashley, 2025) View paper
- Efficiency and Deployment Optimization
  - Quantization and Compression (1 papers)
  - [26] Quamba: A Post-Training Quantization Recipe for Selective State Space Models (Chiang, 2024) View paper
- Representation Learning and Meta-Modeling
  - Weight-Space Representation Learning (2 papers)
  - [13] Towards scalable and versatile weight space learning (SchÃ¼rholt, 2024) View paper
  - [35] Learning useful representations of recurrent neural network weight matrices (Herrmann, 2024) View paper
  - Fast Weight Programming (1 papers)
  - [30] Going beyond linear transformers with recurrent fast weight programmers (Kazuki Irie, 2021) View paper
- Auxiliary Methods and Baselines (3 papers)
  - [42] A Note on Linear Time Series Prediction (Christopher Bonenberger, 2024) View paper
  - [43] Artificial neural network hysteresis operators for the identification of Hammerstein hysteretic systems (Konstantinos Krikelis, 2021) View paper
  - [46] Efficient time-series approximation with linear recurrent neural networks: architecture learning and predictive power (Frieder Stolzenburg, 2025) View paper

Narrative

Core task: sequence modeling with weight-space linear recurrence. The field encompasses a diverse set of approaches that leverage linear recurrence relations—either in activation space or in weight space—to process sequential data efficiently. The taxonomy organizes this landscape into several main branches: Linear Recurrent Architectures and Mechanisms explores core model designs such as state-space models (Mamba[4]), gated variants (Griffin[1], GateLoop[5]), and linear recurrent units (Linear Recurrent Units[2], Behavior-dependent LRU[3]); Training and Optimization Methods addresses how these models are learned; Theoretical Foundations and Analysis investigates expressiveness and convergence properties (Universality Linear Recurrences[34]); Domain-Specific Applications targets tasks like time series forecasting and speech recognition; Efficiency and Deployment Optimization focuses on hardware-aware implementations; Representation Learning and Meta-Modeling examines higher-order abstractions over network parameters (Universal Neural Functionals[40], Scalable Weight Space[13]); and Auxiliary Methods and Baselines provide comparative benchmarks and hybrid designs.

A particularly active line of work centers on architectures that balance expressiveness with computational efficiency, contrasting traditional gated recurrences (Resurrecting RNNs[8]) with newer linear-time mechanisms (Mamba[4], Griffin[1]) and exploring bidirectional processing (Bidirectional Linear Recurrent[9]). Another emerging theme is meta-learning and weight-space recurrence, where models operate on or generate parameters of other networks rather than directly on input sequences. Weight-Space Linear RNN[0] sits squarely within this meta-modeling branch, sharing conceptual ground with MesaNet[20] and Longhorn[24], which also treat network weights as evolving sequences. Compared to these neighbors, Weight-Space Linear RNN[0] emphasizes linear recurrence dynamics in parameter space, offering a distinct angle on how recurrent structure can be embedded at the level of model weights rather than activations. This positioning highlights ongoing questions about where recurrence should be applied—whether in feature representations, in gating mechanisms, or in the parameterization itself—and how such choices affect both learning dynamics and generalization.

Related Works in Same Category

The following **2 sibling papers** share the same taxonomy leaf node with the original paper:

1. MesaNet: Sequence Modeling by Locally Optimal Test-Time Training

Authors: von Oswald, Johannes, Scherrer, Nino, Kobayashi, et al. (24 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

Sequence modeling is currently dominated by causal transformer architectures that use softmax self-attention. Although widely adopted, transformers require scaling memory and compute linearly during inference. A recent stream of work linearized the softmax operation, resulting in powerful recurrent neural network (RNN) models with constant memory and compute costs such as DeltaNet, Mamba or xLSTM. These models can be unified by noting that their recurrent layer dynamics can all be derived from a...

Relationship Analysis

Both papers belong to the Weight-Space and Meta-Learning Recurrence category, exploring sequence modeling through weight-space representations and adaptive learning mechanisms. They overlap in using linear recurrence to update weight-space states (WARP parametrizes hidden states as MLP weights updated via linear dynamics; MesaNet uses optimal test-time regression to learn linear maps in weight-space) and both enable gradient-free adaptation and in-context learning. The key difference is that WARP uses input differences to drive simple linear recurrence ($\theta t = A\theta t\text{-}1 + B\Delta xt$) with the weights directly serving as hidden states, while MesaNet solves an optimal regression problem at each time step using conjugate gradient methods, achieving theoretically optimal solutions but at higher computational cost during inference.

---

2. Longhorn: State Space Models are Amortized Online Learners

Authors: Liu Bo, Wang Rui, Wu Lemeng, Feng, Yihao, et al. (8 authors total) | **Year/Venue**: 2024 • International Conference on Learning Representations | **URL**: View paper

Abstract

Modern large language models are built on sequence modeling via next-token prediction. While the Transformer remains the dominant architecture for sequence modeling, its quadratic decoding complexity in sequence length poses a major limitation. State-space models (SSMs) present a competitive alternative, offering linear decoding efficiency while maintaining parallelism during training. However, most existing SSMs rely on linear recurrence designs that appear somewhat ad hoc. In this work, we exp...

Relationship Analysis

Both papers belong to the Weight-Space and Meta-Learning Recurrence category, exploring linear recurrent architectures for sequence modeling. While the original paper (WARP) explicitly parametrizes hidden states as weights of an auxiliary neural network and uses input differences to drive recurrence, Longhorn frames SSM design through the lens of online learning and derives state transitions from solving online associative recall problems. The key difference is that WARP focuses on weight-space representations with gradient-free adaptation capabilities, whereas Longhorn emphasizes meta-learning principles by conceptualizing SSMs as amortized solutions to online learning objectives.

Contributions Analysis

Overall novelty summary. The paper proposes WARP, a model that parametrizes recurrent hidden states as weights and biases of an auxiliary neural network, driven by input differences through linear recurrence. Within the taxonomy, it resides in the 'Weight-Space and Meta-Learning Recurrence' leaf under 'Linear Recurrent Architectures and Mechanisms'. This leaf contains only three papers total, including WARP itself and two siblings (MesaNet and Longhorn), indicating a relatively sparse and emerging research direction compared to more crowded areas like 'State Space Models' (four papers) or 'Specialized Sequence Tasks' (six papers).

The taxonomy reveals that WARP's neighboring research directions include 'Gated Linear Recurrence Models' (three papers on gating mechanisms), 'State Space Models and Structured Recurrence' (four papers on Mamba-style architectures), and 'Hybrid Linear-Attention Architectures' (three papers combining recurrence with attention). The 'Representation Learning and Meta-Modeling' branch contains related work on weight-space embeddings and fast weight programming, though these focus on learning representations of weights rather than using weights as recurrent states. WARP bridges meta-modeling concepts with linear recurrence, occupying a distinct position at the intersection of these themes.

Among the eleven candidates examined through limited semantic search, no papers were found to clearly refute any of WARP's three contributions. The first contribution (weight-space linear RNN framework) examined one candidate with no refutation. The second contribution (parallelizable training algorithms) examined zero candidates, leaving its novelty unassessed within this search scope. The third contribution (benchmark performance) examined ten candidates, none providing overlapping prior work. This suggests that within the limited search radius, WARP's approach appears relatively distinct, though the small candidate pool (eleven total) means substantial related work may exist beyond this scope.

Based on the limited literature search covering eleven candidates from semantic similarity, WARP appears to occupy a sparsely populated research niche. The taxonomy structure confirms that weight-space recurrence remains an emerging area with few direct comparisons. However, the restricted search scope—examining only top-K semantic matches rather than exhaustive citation networks—means this assessment reflects local novelty within the examined neighborhood rather than comprehensive field coverage. Broader exploration of meta-learning and neural ODE literature could reveal additional relevant prior work.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: Weight-space linear RNN framework with input-difference-driven recurrence

Description: The authors introduce a novel framework that parametrizes RNN hidden states as weights of an auxiliary neural network and uses input differences (rather than raw inputs) to drive linear recurrence. This design combines the efficiency of linear recurrence with the expressivity of non-linear decoding, and is claimed to be the first to treat weight-space features as intermediate hidden state representations in a recurrence.

This contribution was assessed against **1 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Difference between memory and prediction in linear recurrent networks.

URL: View paper

Brief Assessment

The candidate paper (Memory versus Prediction[51]) is not available for comparison. Without access to the full text, it is impossible to assess whether it contains prior work that refutes the novelty of the original paper's weight-space parametrization approach.

Contribution 2: Two parallelizable training algorithms enabling gradient-free adaptation, in-context learning, and physics-informed modeling

Description: The authors present two training modes (convolutional and recurrent) that unlock three practical capabilities: gradient-free adaptation of the auxiliary network at test-time, in-context learning without parameter finetuning, and seamless integration of domain-specific physical priors. A physics-informed variant (WARP-Phys) is shown to achieve an order of magnitude lower error on dynamical system reconstruction tasks.

This contribution was assessed against **0 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

Contribution 3: Extensive real-world benchmark suite demonstrating state-of-the-art performance on multivariate time series classification

Description: The authors establish a comprehensive evaluation suite spanning classification, reconstruction, adaptation, and memory tasks. Their model achieves top-three performance on five out of six challenging multivariate time series classification datasets, demonstrating competitive or superior results compared to established RNNs, state-space models, and Transformers on tasks requiring both short- and long-range dependency modeling.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Unveiling the multi-dimensional spatio-temporal fusion transformer (MDSTFT): A revolutionary deep learning framework for enhanced multi-variate time series â¦

URL: View paper

Brief Assessment

MDSTFT[58] focuses on multivariate time series forecasting with spatio-temporal fusion transformers, not classification benchmarks. The candidate's minimal context does not address classification tasks or benchmark suites comparable to the original paper's UEA evaluation.

2. Knowledge aggregation transformer network for multivariate time series classification

URL: View paper

Brief Assessment

Knowledge Aggregation Transformer[54] focuses on dual-network architectures for MTSC with specific aggregation mechanisms, while the original paper presents a weight-space linear RNN framework. The candidate does not challenge the original's novelty claim of achieving top-three performance on 5 out of 6 challenging datasets using their specific approach.

3. Effectively modeling time series with simple discrete state spaces

URL: View paper

Brief Assessment

Discrete State Spaces[53] focuses on univariate/multivariate time series forecasting and classification using companion matrix SSMs, while the original paper emphasizes weight-space linear RNNs with gradient-free adaptation. The candidate's classification results are on different datasets (ECG, speech audio) than the original's UEA benchmark, making direct novelty comparison difficult.

4. Integration of Mamba and Transformer - MAT for Long-Short Range Time Series Forecasting with Application to Weather Dynamics

URL: View paper

Brief Assessment

MAT[60] focuses on long-short range time series forecasting using weather datasets, not multivariate time series classification tasks. The candidate evaluates forecasting performance with MSE/MAE metrics, while the original contribution emphasizes classification accuracy on UEA datasets.

5. Probabilistic transformer for time series analysis

URL: View paper

Brief Assessment

Probabilistic Transformer[57] focuses on time series forecasting and human motion prediction tasks, not classification. The paper evaluates on forecasting datasets (Solar, Electricity, Traffic, Taxi, Wikipedia) and motion prediction (Human3.6M, HumanEva-I), which are fundamentally different from the classification benchmarks (UEA datasets) used in the original paper.

6. SVP-T: A shape-level variable-position transformer for multivariate time series classification

URL: View paper

Brief Assessment

SVP-T[56] focuses on a shape-level transformer architecture for multivariate time series classification using the UEA archive, while the original paper presents a weight-space linear RNN framework evaluated across diverse tasks including classification, forecasting, and dynamical system reconstruction. The evaluation methodologies and architectural approaches differ fundamentally.

7. Time Series Analysis from Classical Methods to Transformer-Based Approaches: A Review

URL: View paper

Brief Assessment

Time Series Review[55] is a survey paper that reviews existing methods for time series analysis. It does not present original experimental results or claim to achieve state-of-the-art performance on multivariate time series classification benchmarks, and therefore cannot refute the novelty of the original paper's empirical contributions.

8. Multivariate Classification of fMRI Time Series with Fused Window Transformers

URL: View paper

Brief Assessment

Fused Window Transformers[59] focuses on fMRI analysis for gender and disease detection, not general multivariate time series classification benchmarks. The domains and evaluation protocols are fundamentally different.

9. Back to recurrent processing at the crossroad of transformers and state-space models

URL: View paper

Brief Assessment

Recurrent Processing Crossroad[6] focuses on recurrent models and state-space models but does not present multivariate time series classification benchmarks or results that would challenge the original paper's novelty claim about achieving top-three performance on 5 out of 6 UEA datasets.

10. Chimera: Effectively modeling multivariate time series with 2-dimensional state space models

URL: View paper

Brief Assessment

Chimera[52] focuses on 2-dimensional state space models for multivariate time series across multiple tasks (classification, forecasting, anomaly detection), while the original paper presents weight-space linear RNNs. The benchmarks and architectural approaches differ fundamentally.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] Weight-Space Linear Recurrent Neural Networks View paper
- [1] Griffin: Mixing gated linear recurrences with local attention for efficient language models View paper
- [2] Linear recurrent units for sequential recommendation View paper
- [3] Behavior-dependent linear recurrent units for efficient sequential recommendation View paper
- [4] Mamba: Linear-Time Sequence Modeling with Selective State Spaces View paper
- [5] GateLoop: Fully Data-Controlled Linear Recurrence for Sequence Modeling View paper
- [6] Back to recurrent processing at the crossroad of transformers and state-space models View paper
- [7] Resona: Improving Context Copying in Linear Recurrence Models with Retrieval View paper
- [8] Resurrecting recurrent neural networks for long sequences View paper
- [9] Bidirectional linear recurrent models for sequence-level multisource fusion View paper
- [10] SKOLR: Structured Koopman Operator Linear RNN for Time-Series Forecasting View paper
- [11] Hierarchically gated recurrent neural network for sequence modeling View paper
- [12] On the Resurgence of Recurrent Models for Long Sequences--Survey and Research Opportunities in the Transformer Era View paper
- [13] Towards scalable and versatile weight space learning View paper
- [14] Hippo: Recurrent memory with optimal polynomial projections View paper
- [15] Deep symbolic regression for recurrent sequences View paper
- [16] On Structured State-Space Duality View paper
- [17] Test time learning for time series forecasting View paper
- [18] Liger: Linearizing Large Language Models to Gated Recurrent Structures View paper
- [19] Linear-MoE: Linear Sequence Modeling Meets Mixture-of-Experts View paper
- [20] MesaNet: Sequence Modeling by Locally Optimal Test-Time Training View paper
- [21] Input-output HMMs for sequence processing View paper
- [22] Scheduled sampling for sequence prediction with recurrent neural networks View paper
- [23] Evolino: Hybrid neuroevolution/optimal linear search for sequence prediction View paper
- [24] Longhorn: State Space Models are Amortized Online Learners View paper
- [25] An integrated mediapipe-optimized GRU model for Indian sign language recognition View paper
- [26] Quamba: A Post-Training Quantization Recipe for Selective State Space Models View paper
- [27] A TCN-Linear Hybrid Model for Chaotic Time Series Forecasting View paper
- [28] Long short-term memory recurrent neural network architectures for large scale acoustic modeling. View paper
- [29] Mixture-of-Linear-Experts for Long-term Time Series Forecasting View paper
- [30] Going beyond linear transformers with recurrent fast weight programmers View paper
- [31] Regularized dynamic mode decomposition algorithm for time sequence predictions View paper
- [32] Linearly recurrent autoencoder networks for learning dynamics View paper
- [33] GoldFinch: High Performance RWKV/Transformer Hybrid with Linear Pre-Fill and Extreme KV-Cache Compression View paper
- [34] Universality of Linear Recurrences Followed by Non-linear Projections: Finite-Width Guarantees and Benefits of Complex Eigenvalues View paper
- [35] Learning useful representations of recurrent neural network weight matrices View paper
- [36] Unlocking the Secrets of Linear Complexity Sequence Model from A Unified Perspective View paper
- [37] Caudits: Causal disentangled domain adaptation of multivariate time series View paper
- [38] Assessment of unsteady flow predictions using hybrid deep learning based reduced-order models View paper
- [39] V"Mean"ba: Visual State Space Models only need 1 hidden dimension View paper
- [40] Universal neural functionals View paper
- [41] A Financial Time-Series Prediction Model Based on Multiplex Attention and Linear Transformer Structure View paper
- [42] A Note on Linear Time Series Prediction View paper
- [43] Artificial neural network hysteresis operators for the identification of Hammerstein hysteretic systems View paper
- [44] Unifying Linear-Time Attention via Latent Probabilistic Modelling View paper
- [45] Deep Recurrent Neural Networks for Sequential Phenotype Prediction in Genomics View paper
- [46] Efficient time-series approximation with linear recurrent neural networks: architecture learning and predictive power View paper
- [47] LuKAN: A Kolmogorov-Arnold Network Framework for 3D Human Motion Prediction View paper
- [48] Temporal Linear Item-Item Model for Sequential Recommendation View paper

- [49] Enhanced Multi-Vehicle Trajectory Prediction via an Extended Temporal Sequence Fusion Attention Network View paper
- [50] Linear Diffusion Networks View paper
- [51] Difference between memory and prediction in linear recurrent networks. View paper
- [52] Chimera: Effectively modeling multivariate time series with 2-dimensional state space models View paper
- [53] Effectively modeling time series with simple discrete state spaces View paper
- [54] Knowledge aggregation transformer network for multivariate time series classification View paper
- [55] Time Series Analysis from Classical Methods to Transformer-Based Approaches: A Review View paper
- [56] SVP-T: A shape-level variable-position transformer for multivariate time series classification View paper
- [57] Probabilistic transformer for time series analysis View paper
- [58] Unveiling the multi-dimensional spatio-temporal fusion transformer (MDSTFT): A revolutionary deep learning framework for enhanced multi-variate time series â¦ View paper
- [59] Multivariate Classification of fMRI Time Series with Fused Window Transformers View paper
- [60] Integration of Mamba and Transformer - MAT for Long-Short Range Time Series Forecasting with Application to Weather Dynamics View paper