

# Novelty Assessment Report

**Paper:** High-Dimensional Analysis of Single-Layer Attention for Sparse-Token Classification

**PDF URL:** <https://openreview.net/pdf?id=Ae7VWAEIAW>

**Venue:** ICLR 2026 Conference Submission

**Year:** 2026

**Report Generated:** 2025-12-30

## Abstract

When and how can an attention mechanism learn to selectively attend to informative tokens, thereby enabling detection of weak, rare, and sparsely located features? We address these questions theoretically in a sparse-token classification model in which positive samples embed a weak signal vector in a randomly chosen subset of tokens, whereas negative samples are pure noise. For a simple single-layer attention classifier, we show that in the long-sequence limit it can, in principle, achieve vanishing test error when the signal strength grows only logarithmically in the sequence length  $L$ , whereas linear classifiers require  $\sqrt{L}$  scaling. Moving from representational power to learnability, we study training at finite  $L$  in a high-dimensional regime, where sample size and embedding dimension grow proportionally. We prove that just two gradient updates suffice for the query weight vector of the attention classifier to acquire a nontrivial alignment with the hidden signal, inducing an attention map that selectively amplifies informative tokens. We further derive an exact asymptotic expression for the test error of the trained attention-based classifier, and quantify its capacity---the largest dataset size that is typically perfectly separable---thereby explaining the advantage of adaptive token selection over nonadaptive linear baselines.

### Disclaimer

This report is **AI-GENERATED** using Large Language Models and WisPaper (a scholar search engine). It analyzes academic papers' tasks and contributions against retrieved prior work. While this system identifies **POTENTIAL** overlaps and novel directions, **ITS COVERAGE IS NOT EXHAUSTIVE AND JUDGMENTS ARE APPROXIMATE**. These results are intended to assist human reviewers and **SHOULD NOT** be relied upon as a definitive verdict on novelty.

Note that some papers exist in multiple, slightly different versions (e.g., with different titles or URLs). The system may retrieve several versions of the same underlying work. The current automated pipeline does not reliably align or distinguish these cases, so human reviewers will need to disambiguate them manually.

If you have any questions, please contact: mingzhang23@m.fudan.edu.cn

## Core Task Landscape

This paper addresses: **Learning Sparse Weak and Rare Signals in Sequential Data Using Attention Mechanisms**

A total of **50 papers** were analyzed and organized into a taxonomy with **24 categories**.

### Taxonomy Overview

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

- **Theoretical Foundations and Algorithmic Mechanisms**
- **Biomedical and Physiological Signal Processing**
- **Physical Signal Detection and Reconstruction**
- **Multimedia and Vision Applications**
- **Industrial and Mechanical Fault Diagnosis**
- **Recommendation Systems and User Modeling**
- **Spatiotemporal and Event-Based Modeling**
- **Specialized Domain Applications**

### Complete Taxonomy Tree

- Learning Sparse Weak and Rare Signals in Sequential Data Using Attention Mechanisms Survey Taxonomy
- Theoretical Foundations and Algorithmic Mechanisms
  - High-Dimensional Asymptotic Theory ★ (1 papers)
  - [0] High-Dimensional Analysis of Single-Layer Attention for Sparse-Token Classification (Anon et al., 2026) [View paper](#)
  - Sparse Attention Architectures (3 papers)
  - [7] SPARTA: Advancing Sparse Attention in Spiking Neural Networks via Spike-Timing-Based Prioritization (Kim Changick, 2025) [View paper](#)
  - [43] Sparse and Continuous Attention Mechanisms (Martins, 2022) [View paper](#)
  - [48] Block Sparse Flash Attention (Daniel Ohayon, 2025) [View paper](#)
  - Hybrid Attention Mechanisms (4 papers)
  - [8] Sia-net: Sparse interactive attention network for multimodal emotion recognition (Shuzhen Li, 2024) [View paper](#)
  - [10] Reinforced self-attention network: a hybrid of hard and soft attention for sequence modeling (Shen Tao, 2018) [View paper](#)
  - [12] Generalized uncertainty-based evidential fusion with hybrid multi-head attention for weak-supervised temporal action localization (Yuanpeng He, 2024) [View paper](#)
  - [31] A hybrid attention mechanism for weakly-supervised temporal action localization (Ashrafal Islam, 2021) [View paper](#)
- Biomedical and Physiological Signal Processing
  - Neurophysiological Signal Analysis (3 papers)
  - [3] CSBrain: A Cross-scale Spatiotemporal Brain Foundation Model for EEG Decoding (Zhou Yuchen, 2025) [View paper](#)
  - [13] EEG anomaly detection using temporal graph attention for clinical applications (Xiao-pei, 2025) [View paper](#)
  - [16] STDCformer: Spatial-temporal dual-path cross-attention model for fMRI-based autism spectrum disorder identification (Haifeng Zhang, 2024) [View paper](#)
  - Cardiac Signal Processing (1 papers)
  - [24] Deep Learning-Based Data Augmentation and Model Fusion for Automatic Arrhythmia Identification and Classification Algorithms (Shuai Ma, 2022) [View paper](#)
  - Remote Physiological Monitoring (2 papers)
  - [5] Emotion Recognition from rPPG via Physiologically Inspired Temporal Encoding and Attention-Based Curriculum Learning (Changmin Lee, 2025) [View paper](#)

- [45] Through-Wall Human Mesh Recovery Using Radio Signals (M. Zhao, 2019) [View paper](#)
- Clinical Prediction and Diagnosis (2 papers)
- [11] MEET-Sepsis: Multi-Endogenous-View Enhanced Time-Series Representation Learning for Early Sepsis Prediction (Tan Ze-xi, 2025) [View paper](#)
- [19] Clinical NLP with Attention-Based Deep Learning for Multi-Disease Prediction (Ting Xu, 2025) [View paper](#)
- Single-Cell and Molecular Data (1 papers)
- [39] Explainable modeling of single-cell perturbation data using attention and sparse dictionary learning. (Xu Yang, 2025) [View paper](#)
- Physical Signal Detection and Reconstruction
  - Seismic and Geophysical Signal Recovery (2 papers)
  - [23] Complete perception self-attention network for weak seismic signal recovery in distributed acoustic sensing vertical seismic profile data (Jilei Sui, 2023) [View paper](#)
  - [25] Acoustic impedance prediction using an attention-based dual-branch double-inversion network (Wen Feng, 2025) [View paper](#)
  - Acoustic and Radar Signal Processing (3 papers)
  - [14] Source identification of weak audio signals using attention based convolutional neural network: K. Presannakumar and A. Mohamed (K Presannakumar, 2023) [View paper](#)
  - [15] Attention-Refined Unrolling for Sparse Sequential Micro-Doppler Reconstruction (Riccardo Mazzieri, 2024) [View paper](#)
  - [21] A spatiotemporal information-driven cross-attention model with sparse representation for gnss nlos signal classification (Kungan Zeng, 2024) [View paper](#)
  - Chaotic and Noisy Signal Detection (1 papers)
  - [2] Weak signal detection in chaotic noise background-based on VMD-EEMD and self-attention mechanisms (Shengli Zhao, 2024) [View paper](#)
  - Compressive Sensing and Reconstruction (2 papers)
  - [30] Transformer end-to-end optimization of compressive LiDARs using imaging spectroscopy side information (Nestor Porras-Diaz, 2024) [View paper](#)
  - [36] Ufc-net: Unrolling fixed-point continuous network for deep compressive sensing (Xiaoyang Wang, 2024) [View paper](#)
- Multimedia and Vision Applications
  - Video Captioning and Understanding (2 papers)
  - [1] Swinbert: End-to-end transformers with sparse attention for video captioning (Kevin Lin, 2022) [View paper](#)
  - [32] Svitt: Temporal learning of sparse video-text transformers (Yi Li, 2023) [View paper](#)
  - Temporal Action Localization (2 papers)
  - [34] Weakly supervised action localization by sparse temporal pooling network (Nguyen, 2018) [View paper](#)
  - [37] Sequential attention mechanism for weakly supervised video anomaly detection (Waseem Ullah, 2023) [View paper](#)
  - Small Target and Infrared Detection (1 papers)
  - [6] An Improved U-Net Infrared Small Target Detection Algorithm Based on Multi-Scale Feature Decomposition and Fusion and Attention Mechanism (Xiangsuo Fan, 2024) [View paper](#)
  - Multimodal and Cross-Modal Fusion (2 papers)
  - [26] HunyuanVideo-HOMA: Generic Human-Object Interaction in Multimodal Driven Human Animation (Huang Ziyao, 2025) [View paper](#)
  - [35] Top-Down Cross-Modal Guidance for Robust RGB-T Tracking (Liang Chen, 2024) [View paper](#)
- Industrial and Mechanical Fault Diagnosis (3 papers)
  - [4] Rolling bearing fault diagnosis method base on periodic sparse attention and LSTM (Yiyao An, 2022) [View paper](#)
  - [20] Sparse Attention Coding-Timestep Shrinking Spiking Neural Network for Gearbox Fault Diagnosis (Shilong Zhu, 2025) [View paper](#)
  - [38] Temporal-Spatial Attention Network: A Novel Axial Piston Pump Coupled Fault Diagnosis Method (Shihao Liu, 2024) [View paper](#)
- Recommendation Systems and User Modeling
  - Long-Sequence and Multimodal Recommendation (3 papers)
  - [9] UMI-Rec: A Unified Multi-modal Intent Fusion Framework with State-Space Models and Large Language Models for Recommendation (Zare, 2025) [View paper](#)
  - [41] HeteroTempRec: Temporally-aware heterogeneous architecture with sparse block attention for efficient sequential recommendation (Zihao Song, 2025) [View paper](#)
  - [50] Multimodal Fusion And Sparse Attention-based Alignment Model for Long Sequential Recommendation (Fu YongRui, 2025) [View paper](#)
  - Graph-Based and Contrastive Recommendation (1 papers)
  - [47] Contrastive Learning of Sequential Recommendation with Graph Attention Mechanisms (Xu Xie, 2024) [View paper](#)
  - Real-Time Advertising and Dynamic Interest (1 papers)
  - [33] Multi-Level Attention and Sequence Modeling for Dynamic User Interest Representation in Real-Time Advertising Recommendation (Ming, 2023) [View paper](#)
- Spatiotemporal and Event-Based Modeling
  - Spatiotemporal Data Inference (3 papers)
  - [27] Spatial-temporal data inference with graph attention neural networks in sparse mobile crowdsensing (Guisong Yang, 2024) [View paper](#)
  - [28] Interpretable Traffic Accident Prediction: Attention Spatial-Temporal Multi-Graph Traffic Stream Learning Approach (Chaojie Li, 2024) [View paper](#)
  - [42] A static-dynamic feature coupling framework for spatiotemporal multiphase flow prediction in heterogeneous reservoirs (W Chen, 2025) [View paper](#)
  - Event Stream Classification (1 papers)
  - [18] Temporal-wise attention spiking neural networks for event streams classification (Man Yao, 2021) [View paper](#)
  - Anomaly Detection and Rare Event Prediction (2 papers)
  - [40] Prediction of Anomalous Events With Data Augmentation and Hybrid Deep Learning Approach (Ahmed Shoyeb Raihan, 2024) [View paper](#)
  - [49] Attention-Based Dynamic Sampling for AI-Optimized Smart Grids: A Unified Framework for Robust Load Balancing and Renewable Integration (Shanguo Zhao, 2025) [View paper](#)

- Specialized Domain Applications (5 papers)
  - [17] Heterogeneous convolutional recurrent neural network with attention mechanism and feature aggregation for voice activity detection (Yingwei Tan, 2024) [View paper](#)
  - [22] Mixture-of-experts graph transformers for interpretable particle collision detection (Donatella Genovese, 2025) [View paper](#)
  - [29] Sequential weakly labeled multiactivity localization and recognition on wearable sensors using recurrent attention networks (Kun Wang, 2021) [View paper](#)
  - [44] Visual attention emerges from recurrent sparse reconstruction (Shi, 2022) [View paper](#)
  - [46] A sparse-based transformer network with associated spatiotemporal feature for micro-expression recognition (Jie Zhu, 2022) [View paper](#)

## Narrative

Core task: learning sparse weak and rare signals in sequential data using attention mechanisms. The field spans a diverse set of branches, each addressing distinct problem settings and methodological emphases. Theoretical Foundations and Algorithmic Mechanisms explores the mathematical underpinnings and novel attention architectures, including sparse attention variants and high-dimensional asymptotic analyses. Biomedical and Physiological Signal Processing focuses on extracting subtle patterns from clinical time series, such as EEG anomalies and cardiac arrhythmias, often leveraging domain-specific preprocessing. Physical Signal Detection and Reconstruction tackles radar, seismic, and acoustic data where weak targets or events must be isolated from noise. Multimedia and Vision Applications address video understanding, action localization, and micro-expression recognition, where rare frames or fleeting cues carry critical information. Industrial and Mechanical Fault Diagnosis applies attention to vibration and sensor streams for early fault detection. Recommendation Systems and User Modeling capture sparse user interactions and evolving preferences over time. Spatiotemporal and Event-Based Modeling handles irregular or event-driven data, while Specialized Domain Applications cover niche areas like particle collision detection and smart grid sampling.

Several active lines of work highlight contrasting trade-offs between computational efficiency and expressive power. Sparse attention mechanisms such as Periodic Sparse Attention[4] and Block Sparse Flash[48] reduce quadratic complexity, enabling longer context windows, whereas dense architectures like Hybrid Attention Mechanism[31] prioritize richer feature interactions at higher cost. In biomedical domains, works like CSBrain[3] and rPPG Emotion Recognition[5] emphasize multimodal fusion and physiological priors, while industrial applications such as Gearbox Fault Diagnosis[20] rely on signal decomposition and temporal pooling. The original paper, Single-Layer Attention[0], resides within the Theoretical Foundations branch under High-Dimensional Asymptotic Theory, offering a rigorous analysis of how single-layer attention behaves in high-dimensional regimes. This positions it as a foundational complement to empirical studies like Reinforced Self-Attention[10] and Sparse Continuous Attention[43], which explore architectural innovations without the same level of theoretical grounding. Open questions remain around scaling these insights to deeper networks and bridging theory with domain-specific constraints seen in applied branches.

## Related Works in Same Category

No sibling papers were found in the same taxonomy leaf. A taxonomy-subtopic-level comparison will be produced instead.

### Taxonomy-Level Summary

The original leaf focuses on theoretical foundations—proving convergence guarantees and capacity bounds for attention mechanisms in high-dimensional settings. Its siblings address practical design: Hybrid Attention Mechanisms combines multiple attention types or modalities, while Sparse Attention Architectures develops efficient sparse patterns for implementation. Together, they span the theory-to-practice spectrum for attention-based learning of weak signals.

**Similarities:** - All three subtopics concern attention mechanisms applied to sequential or structured data - Each addresses scalability challenges: the original leaf through asymptotic theory, siblings through architectural innovation - All aim to improve signal extraction—whether through theoretical guarantees (original) or computational efficiency (siblings)

**Differences:** - Original leaf emphasizes provable guarantees and mathematical analysis; siblings prioritize architectural design and empirical performance - Hybrid Attention Mechanisms integrates diverse attention types or modalities; original leaf analyzes single-mechanism properties in high dimensions - Sparse Attention Architectures focuses on implementation-level sparsity patterns (block-sparse, entmax); original leaf studies theoretical capacity and convergence without implementation constraints - Original leaf excludes empirical-only work; siblings exclude purely theoretical studies, creating complementary scopes

**Suggested Search Directions:** - Bridging work that validates high-dimensional theoretical predictions on sparse attention architectures - Asymptotic analysis of hybrid attention systems combining multiple mechanisms - Sample complexity bounds for resource-adaptive or entmax-based sparse attention

### Sibling Subtopics

- **Hybrid Attention Mechanisms** (leaves: 1, papers: 4)
  - Scope: Integration of multiple attention types (hard/soft, multi-head, cross-modal) or fusion with other learning paradigms.
  - Exclude: Excludes single-modality or single-attention-type methods; see Sparse Attention Architectures or domain applications.
- **Sparse Attention Architectures** (leaves: 1, papers: 3)
  - Scope: Novel sparse attention designs including block-sparse, entmax, or resource-adaptive mechanisms for computational efficiency.
  - Exclude: Excludes theoretical analysis without implementation; see High-Dimensional Asymptotic Theory.

## Contributions Analysis

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

### Contribution 1: Exponential separation in signal strength requirements between attention and linear classifiers

**Description:** The authors prove that in the limit of large sequence length, attention models can detect signals that are exponentially weaker than those detectable by linear classifiers. Specifically, attention requires signal strength  $\theta = \log L$  while linear classifiers need  $\theta = \sqrt{L}$  for perfect classification.

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. Radial Attention: $\mathcal{O}(n \log n)$ Sparse Attention with Energy Decay for Long Video Generation

URL: [View paper](#)

##### Brief Assessment

Radial Attention[51] focuses on sparse attention mechanisms for video generation with spatiotemporal energy decay, not on theoretical signal detection scaling requirements in classification tasks. The candidate addresses computational efficiency in video diffusion models rather than statistical learning theory comparisons between attention and linear classifiers.

## Contribution 2: Exact asymptotic characterization of test error after two gradient updates

**Description:** The authors derive an exact asymptotic expression for the test error of the attention classifier after only two gradient steps on the query weights, followed by full optimization of readout weights. This characterization is precise down to explicit constants in a high-dimensional regime where sample size and embedding dimension grow proportionally.

This contribution was assessed against **7 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

### 1. One step of gradient descent is provably the optimal in-context learner with one layer of linear self-attention

URL: [View paper](#)

#### Brief Assessment

Gradient Descent Learner[61] analyzes one-layer linear self-attention transformers trained on linear regression, proving the global minimum implements one gradient step. The original paper studies attention classifiers on sparse-token classification with two gradient updates followed by full optimization of readout weights, providing exact asymptotic test error characterization in a different high-dimensional regime.

---

### 2. On the role of attention in prompt-tuning

URL: [View paper](#)

#### Brief Assessment

Prompt-Tuning Attention[59] focuses on prompt-tuning in attention architectures with a contextual data model, not on general test error characterization after few gradient updates in attention models for sparse-token classification.

---

### 3. Benign overfitting in single-head attention

URL: [View paper](#)

#### Brief Assessment

Benign Overfitting[60] focuses on benign overfitting in single-head attention with label-flipping noise, analyzing when interpolation occurs without harming generalization. The original paper characterizes test error after two gradient steps in a sparse-token classification task without label noise, using a different data distribution and model architecture.

---

### 4. Incremental few-shot learning with attention attractor networks

URL: [View paper](#)

#### Brief Assessment

Attention Attractor Networks[62] focuses on incremental few-shot learning with attention mechanisms for class-incremental scenarios, not on deriving exact asymptotic test error characterizations after gradient updates in high-dimensional attention models.

---

### 5. Leveraging task variability in meta-learning

URL: [View paper](#)

#### Brief Assessment

Task Variability Meta-Learning[63] focuses on meta-learning across task distributions with attention networks, not on deriving exact asymptotic test error expressions after few gradient steps in sparse-token classification models.

---

### 6. Meta-Learning for Adaptive Dynamical System Characterization

URL: [View paper](#)

#### Brief Assessment

Adaptive Dynamical System[58] focuses on meta-learning for dynamical system characterization with causal attention masking, not on deriving exact asymptotic test error expressions after few gradient steps in attention classifiers for sparse-token tasks.

---

### 7. Transformers learn to implement multi-step gradient descent with chain of thought

URL: [View paper](#)

#### Brief Assessment

Chain of Thought[57] focuses on training transformers with chain-of-thought prompting for in-context learning tasks, not on characterizing test error after gradient updates in attention models for sparse-token classification.

---

## Contribution 3: Capacity characterization quantifying advantage of adaptive token selection

**Description:** The authors characterize the capacity of the attention model, defined as the maximal dataset size that can be perfectly fit with high probability, and compare it with linear classifiers. This provides a complementary perspective on how attention's adaptive token selection mechanism outperforms nonadaptive approaches.

This contribution was assessed against **5 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

### 1. TokenFLEX: Unified VLM Training for Flexible Visual Tokens Inference

URL: [View paper](#)

#### Brief Assessment

TokenFLEX[52] focuses on adaptive visual token allocation in vision-language models for computational efficiency, not on theoretical capacity analysis of attention mechanisms versus linear classifiers in sparse-token classification tasks.

---

### 2. Static or Dynamic: Towards Query-Adaptive Token Selection for Video Question Answering

URL: [View paper](#)

#### Brief Assessment

Query-Adaptive Token Selection[53] focuses on video question answering with adaptive visual token selection based on query requirements, not on theoretical capacity analysis of attention mechanisms versus linear classifiers in sparse-token classification.

---

### 3. AdaptiVision: A Flexible and Efficient Vision Transformer for Adaptive Token Pruning

URL: [View paper](#)

#### Brief Assessment

AdaptiVision[56] focuses on adaptive token pruning in vision transformers for computational efficiency, not on capacity analysis of attention mechanisms versus linear classifiers in sparse-token classification tasks.

---

#### 4. Adaptive Facet Selection in Multidimensional Hosting Capacity Region Assessment

URL: [View paper](#)

##### Brief Assessment

Adaptive Facet Selection[55] addresses hosting capacity region assessment in power grids using adaptive facet selection measures, not attention mechanisms or token selection in machine learning models. The domains are entirely different.

#### 5. Token-based adaptive time-series prediction by ensembling linear and non-linear estimators: a machine learning approach for predictive analytics on big stock data

URL: [View paper](#)

##### Brief Assessment

Token-based Time-Series[54] focuses on ensemble methods for financial time-series prediction using Kalman filters and LSTM networks, not on theoretical capacity analysis of attention mechanisms or adaptive token selection in classification tasks.

### Appendix: Text Similarity Detection

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

### References

- [0] High-Dimensional Analysis of Single-Layer Attention for Sparse-Token Classification [View paper](#)
- [1] Swinbert: End-to-end transformers with sparse attention for video captioning [View paper](#)
- [2] Weak signal detection in chaotic noise background-based on VMD-EEMD and self-attention mechanisms [View paper](#)
- [3] CSBrain: A Cross-scale Spatiotemporal Brain Foundation Model for EEG Decoding [View paper](#)
- [4] Rolling bearing fault diagnosis method base on periodic sparse attention and LSTM [View paper](#)
- [5] Emotion Recognition from rPPG via Physiologically Inspired Temporal Encoding and Attention-Based Curriculum Learning [View paper](#)
- [6] An Improved U-Net Infrared Small Target Detection Algorithm Based on Multi-Scale Feature Decomposition and Fusion and Attention Mechanism [View paper](#)
- [7] SPARTA: Advancing Sparse Attention in Spiking Neural Networks via Spike-Timing-Based Prioritization [View paper](#)
- [8] Sia-net: Sparse interactive attention network for multimodal emotion recognition [View paper](#)
- [9] UMI-Rec: A Unified Multi-modal Intent Fusion Framework with State-Space Models and Large Language Models for Recommendation [View paper](#)
- [10] Reinforced self-attention network: a hybrid of hard and soft attention for sequence modeling [View paper](#)
- [11] MEET-Sepsis: Multi-Endogenous-View Enhanced Time-Series Representation Learning for Early Sepsis Prediction [View paper](#)
- [12] Generalized uncertainty-based evidential fusion with hybrid multi-head attention for weak-supervised temporal action localization [View paper](#)
- [13] EEG anomaly detection using temporal graph attention for clinical applications [View paper](#)
- [14] Source identification of weak audio signals using attention based convolutional neural network: K. Presannakumar and A. Mohamed [View paper](#)
- [15] Attention-Refined Unrolling for Sparse Sequential Micro-Doppler Reconstruction [View paper](#)
- [16] STDCformer: Spatial-temporal dual-path cross-attention model for fMRI-based autism spectrum disorder identification [View paper](#)
- [17] Heterogeneous convolutional recurrent neural network with attention mechanism and feature aggregation for voice activity detection [View paper](#)
- [18] Temporal-wise attention spiking neural networks for event streams classification [View paper](#)
- [19] Clinical NLP with Attention-Based Deep Learning for Multi-Disease Prediction [View paper](#)
- [20] Sparse Attention Coding-Timestep Shrinking Spiking Neural Network for Gearbox Fault Diagnosis [View paper](#)
- [21] A spatiotemporal information-driven cross-attention model with sparse representation for gnss nlos signal classification [View paper](#)
- [22] Mixture-of-experts graph transformers for interpretable particle collision detection [View paper](#)
- [23] Complete perception self-attention network for weak seismic signal recovery in distributed acoustic sensing vertical seismic profile data [View paper](#)
- [24] Deep Learning-Based Data Augmentation and Model Fusion for Automatic Arrhythmia Identification and Classification Algorithms [View paper](#)
- [25] Acoustic impedance prediction using an attention-based dual-branch double-inversion network [View paper](#)
- [26] HunyuanVideo-HOMA: Generic Human-Object Interaction in Multimodal Driven Human Animation [View paper](#)
- [27] Spatial-temporal data inference with graph attention neural networks in sparse mobile crowdsensing [View paper](#)
- [28] Interpretable Traffic Accident Prediction: Attention Spatial-Temporal Multi-Graph Traffic Stream Learning Approach [View paper](#)
- [29] Sequential weakly labeled multiactivity localization and recognition on wearable sensors using recurrent attention networks [View paper](#)
- [30] Transformer end-to-end optimization of compressive LiDARs using imaging spectroscopy side information [View paper](#)
- [31] A hybrid attention mechanism for weakly-supervised temporal action localization [View paper](#)
- [32] Svitt: Temporal learning of sparse video-text transformers [View paper](#)
- [33] Multi-Level Attention and Sequence Modeling for Dynamic User Interest Representation in Real-Time Advertising Recommendation [View paper](#)
- [34] Weakly supervised action localization by sparse temporal pooling network [View paper](#)
- [35] Top-Down Cross-Modal Guidance for Robust RGB-T Tracking [View paper](#)
- [36] Ufc-net: Unrolling fixed-point continuous network for deep compressive sensing [View paper](#)
- [37] Sequential attention mechanism for weakly supervised video anomaly detection [View paper](#)
- [38] Temporal-Spatial Attention Network: A Novel Axial Piston Pump Coupled Fault Diagnosis Method [View paper](#)
- [39] Explainable modeling of single-cell perturbation data using attention and sparse dictionary learning. [View paper](#)
- [40] Prediction of Anomalous Events With Data Augmentation and Hybrid Deep Learning Approach [View paper](#)
- [41] HeteroTempRec: Temporally-aware heterogeneous architecture with sparse block attention for efficient sequential recommendation [View paper](#)
- [42] A static-dynamic feature coupling framework for spatiotemporal multiphase flow prediction in heterogeneous reservoirs [View paper](#)

- [43] Sparse and Continuous Attention Mechanisms [View paper](#)
- [44] Visual attention emerges from recurrent sparse reconstruction [View paper](#)
- [45] Through-Wall Human Mesh Recovery Using Radio Signals [View paper](#)
- [46] A sparse-based transformer network with associated spatiotemporal feature for micro-expression recognition [View paper](#)
- [47] Contrastive Learning of Sequential Recommendation with Graph Attention Mechanisms [View paper](#)
- [48] Block Sparse Flash Attention [View paper](#)
- [49] Attention-Based Dynamic Sampling for AI-Optimized Smart Grids: A Unified Framework for Robust Load Balancing and Renewable Integration [View paper](#)
- [50] Multimodal Fusion And Sparse Attention-based Alignment Model for Long Sequential Recommendation [View paper](#)
- [51] Radial Attention:  $O(n \log n)$  Sparse Attention with Energy Decay for Long Video Generation [View paper](#)
- [52] TokenFLEX: Unified VLM Training for Flexible Visual Tokens Inference [View paper](#)
- [53] Static or Dynamic: Towards Query-Adaptive Token Selection for Video Question Answering [View paper](#)
- [54] Token-based adaptive time-series prediction by ensembling linear and non-linear estimators: a machine learning approach for predictive analytics on big stock data [View paper](#)
- [55] Adaptive Facet Selection in Multidimensional Hosting Capacity Region Assessment [View paper](#)
- [56] AdaptiVision: A Flexible and Efficient Vision Transformer for Adaptive Token Pruning [View paper](#)
- [57] Transformers learn to implement multi-step gradient descent with chain of thought [View paper](#)
- [58] Meta-Learning for Adaptive Dynamical System Characterization [View paper](#)
- [59] On the role of attention in prompt-tuning [View paper](#)
- [60] Benign overfitting in single-head attention [View paper](#)
- [61] One step of gradient descent is provably the optimal in-context learner with one layer of linear self-attention [View paper](#)
- [62] Incremental few-shot learning with attention attractor networks [View paper](#)
- [63] Leveraging task variability in meta-learning [View paper](#)