Novelty Assessment Report

Paper: Temporal Generalization: A Reality Check
PDF URL: https://openreview.net/pdf?id=Wz0ILlbh9U
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2025-12-27

Abstract

Machine learning (ML) models often struggle to maintain performance under distribution shifts, leading to inaccurate predictions on unseen future data. In this work, we investigate whether and under what conditions models can achieve such a generalization when relying solely on past data. We explore two primary approaches: convex combinations of past model parameters (parameter interpolation) and explicit extrapolation beyond the convex hull of past parameters (parameter extrapolation). We benchmark several methods within these categories on a diverse set of temporal tasks, including language modeling, news summarization, news tag prediction, academic paper categorization, satellite image-based land use classification over time, and historical yearbook photo gender prediction. Our empirical findings show that none of the evaluated methods consistently outperforms the simple baseline of using the latest available model parameters in all scenarios. In the absence of access to future data or robust assumptions about the underlying data-generating process, these results underscore the inherent difficulties of generalizing and extrapolating to future data and warrant caution when evaluating claims of such generalization.

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: **Temporal Generalization of Machine Learning Models Under Distribution Shifts**

A total of **50 papers** were analyzed and organized into a taxonomy with **18 categories**.

Taxonomy Overview

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

- **Test-Time Adaptation and Continual Learning**
- **Domain Adaptation and Generalization**
- **Time Series and Temporal Data Methods**
- **Theoretical Foundations and Benchmarking**
- **Application-Specific Methods**

Complete Taxonomy Tree

- Temporal Generalization of Machine Learning Models Under Distribution Shifts Survey Taxonomy
- Test-Time Adaptation and Continual Learning
  - Online Test-Time Adaptation (3 papers)
  - [6] Bayestta: Continual-temporal test-time adaptation for vision-language models via gaussian discriminant analysis (Cui Shuang, 2025) View paper
  - [8] Online test-time adaptation for better generalization of interatomic potentials to out-of-distribution data (Taoyong Cui, 2025) View paper
  - [33] Test-time adaptation to distribution shift by confidence maximization and input transformation (Mummadi Chaithanya Kumar, 2021) View paper
  - Continual Learning Frameworks (2 papers)
  - [4] Adaptive State Estimation and Continual Learning under Data Distribution Shift (Arvin Hosseinzadeh, 2025) View paper
  - [19] Adaptive Personalized Federated Learning for Non-IID Data with Continual Distribution Shift (Sisi Chen, 2024) View paper
  - Source-Free Domain Adaptation (3 papers)
  - [10] Adaptable: Test-time adaptation for tabular data via shift-aware uncertainty calibrator and label distribution handler (Kim Changhun, 2024) View paper
  - [12] Temporal Restoration and Spatial Rewiring for Source-Free Multivariate Time Series Domain Adaptation (Wang Yu-cheng, 2025) View paper
  - [46] Source-free domain adaptation with temporal imputation for time series data (Mohamed Ragab, 2023) View paper
- Domain Adaptation and Generalization
  - Domain Generalization (3 papers)
  - [17] Predicting Practically? Domain Generalization for Predictive Analytics in Real-world Environments (Duan Hanyu, 2025) View paper
  - [26] Universal Domain Adaptation for Robust Handling of Distributional Shifts in NLP (Cho, 2023) View paper
  - [28] Evaluation of domain generalization and adaptation on improving model robustness to temporal dataset shift in clinical medicine (Lin Lawrence Guo, 2021) View paper
  - Unsupervised Domain Adaptation (3 papers)
  - [1] CERA: A Framework for Improved Generalization of Machine Learning Models to Changed Climates (Liu Shuchang, 2025) View paper
  - [11] Improving the Generalization of Segmentation Foundation Model under Distribution Shift via Weakly Supervised Adaptation (Haojie Zhang, 2023) View paper

- [45] CFTResNet: A novel cross-domain diagnosis framework guided by interpretability for cardiovascular diseases. (Pengfei Liang, 2025) View paper
- Multi-Domain and Cross-Domain Learning (2 papers)
- [13] A Multi-modal Architecture with Spatio-Temporal-Text Adaptation for Video-based Traffic Accident Anticipation (Patrik Patera, 2025) View paper
- [47] Monitoring of Laser Powder Bed Fusion process by bridging dissimilar process maps using deep learning-based domain adaptation on acoustic emissions (Richter, 2024) View paper

- Time Series and Temporal Data Methods
  - Time Series Forecasting Under Shift (6 papers)
  - [2] Reversible instance normalization for accurate time-series forecasting against distribution shift (T Kim, 2021) View paper
  - [9] Calibration of time-series forecasting: Detecting and adapting context-driven distribution shift (Mouxiang Chen, 2024) View paper
  - [21] Learning pattern-specific experts for time series forecasting under patch-level distribution shift (Sun YanRu, 2024) View paper
  - [29] Robust Multivariate Time Series Forecasting Against Intraseries and Interseries Transitional Shift (Hui He, 2025) View paper
  - [34] Adaptive temporal transformer method for short-term wind power forecasting considering shift in time series distribution (Dan Li, 2024) View paper
  - [37] Distributional drift adaptation with temporal conditional variational autoencoder for multivariate time series forecasting (Hui He, 2024) View paper
  - Time Series Classification and Detection (2 papers)
  - [3] Diversify: A general framework for time series out-of-distribution detection and generalization (Wang Lu, 2024) View paper
  - [22] Towards Streaming Land Use Classification of Images with Temporal Distribution Shifts (Lorenzo Iovine, 2025) View paper
  - Graph-Based Temporal Learning (4 papers)
  - [7] Dynamic graph neural networks under spatio-temporal distribution shift (Z Zhang, 2022) View paper
  - [18] Improving Generalization of Dynamic Graph Learning via Environment Prompt (Qihe Huang, 2024) View paper
  - [27] Graph learning under distribution shifts: A comprehensive survey on domain adaptation, out-of-distribution, and continual learning (Wu Man, 2024) View paper
  - [32] A survey of deep graph learning under distribution shifts: from graph out-of-distribution generalization to adaptation (Zhang Ke-xin, 2024) View paper
- Theoretical Foundations and Benchmarking
  - Theoretical Analysis and Estimation (3 papers)
  - [14] Adaptive Conformal Inference Under Distribution Shift (Gibbs, 2021) View paper
  - [39] Efficient Non-stationary Online Learning by Wavelets with Applications to Online Distribution Shift Adaptation (Yuyang Qian, 2024) View paper
  - [49] Adaptive Estimation and Learning under Temporal Distribution Shift (Baby, 2025) View paper
  - Benchmarking and Evaluation ★ (4 papers)
  - [0] Temporal Generalization: A Reality Check (Anon et al., 2026) View paper
  - [20] Wild-time: A benchmark of in-the-wild distribution shift over time (Yao, 2022) View paper
  - [30] Understanding the Limits of Deep Tabular Methods with Temporal Shift (Ye, 2025) View paper
  - [36] Out-of-Distribution Generalization in Time Series: A Survey (Wu Xin, 2025) View paper
  - Model Selection and Assessment (2 papers)
  - [24] Model Assessment and Selection under Temporal Distribution Shift (Wang, 2024) View paper
  - [40] Temporal Evaluation ofÂ Uncertainty Quantification Under Distribution Shift (Emma Svensson, 2024) View paper
- Application-Specific Methods
  - Healthcare and Clinical Applications (2 papers)
  - [23] Temporal Distribution Shift in Real-World Pharmaceutical Data: Implications for Uncertainty Quantification in QSAR Models (Hannah Rosa Friesacher, 2025) View paper
  - [25] EHR foundation models improve robustness in the presence of temporal distribution shift (Lin Lawrence Guo, 2023) View paper
  - Recommender Systems and User Modeling (1 papers)
  - [5] Generalizable Recommender System During Temporal Popularity Distribution Shifts (Hyunsik Yoo, 2025) View paper
  - Natural Language Processing Applications (3 papers)
  - [15] Simple temporal adaptation to changing label sets: Hashtag prediction via dense KNN (Niloofar Mireshghallah, 2023) View paper
  - [31] Learning to Align: Addressing Character Frequency Distribution Shifts in Handwritten Text Recognition (Pavlopoulos, 2025) View paper
  - [48] Assessing and Mitigating Medical Knowledge Drift and Conflicts in Large Language Models (Wu Weiyi, 2025) View paper
  - Computer Vision Applications (1 papers)
  - [38] Dual prototype evolving for test-time generalization of vision-language models (Simon Stepputtis, 2024) View paper
  - Tabular Data and Structured Prediction (1 papers)
  - [16] Drift-Resilient TabPFN: In-Context Learning Temporal Distribution Shifts on Tabular Data (Kai Helli, 2024) View paper
  - Specialized Domain Applications (6 papers)
  - [35] Robust Machine Learning: Detection, Evaluation and Adaptation Under Distribution Shift (Garg, 2024) View paper
  - [41] Asp: Learn a universal neural solver! (Chenguang Wang, 2024) View paper
  - [42] Handling new class in online label shift (Yu-Yang Qian, 2025) View paper
  - [43] Adaptive streaming architectures: Integrating change-data-capture with online machine learning for real-time enterprise decisioning (Kumar, 2025) View paper
  - [44] Open Spatio-Temporal Foundation Models for Traffic Prediction (Zhonghang Li, 2024) View paper
  - [50] Deep Learning-Based Financial Fraud Detection with Temporal and Feature-Level Adaptation (Tsai, 2025) View paper

Narrative

Core task: temporal generalization of machine learning models under distribution shifts. The field addresses how models maintain performance when data distributions evolve over time, a challenge spanning diverse application domains and methodological traditions. The taxonomy organizes this landscape into five main branches. Test-Time Adaptation and Continual Learning focuses on methods that update models dynamically as new data arrives, often without access to original training distributions. Domain Adaptation and Generalization emphasizes learning representations that transfer across different but related distributions, including techniques for aligning source and target domains. Time Series and Temporal Data Methods develops specialized architectures and algorithms for sequential data where temporal dependencies are explicit. Theoretical Foundations and Benchmarking provides the mathematical underpinnings and standardized evaluation protocols needed to compare approaches rigorously, as seen in works like Wild-time Benchmark[20] and Time Series OOD Survey[36]. Application-Specific Methods tailors solutions to particular domains such as healthcare, climate science, or finance, where domain constraints shape the nature of temporal shifts.

Several active research directions reveal key trade-offs in the field. One tension involves the balance between adaptation speed and stability: methods like Bayestta Continual Adaptation[6] and Adaptive Conformal Inference[14] must update quickly to track shifts while avoiding catastrophic forgetting or overfitting to transient noise. Another contrast emerges between model-centric approaches that modify architectures or training procedures versus data-centric methods that characterize and correct for specific shift patterns, as explored in works addressing label shift, covariate shift, and concept drift. Temporal Generalization Reality Check[0] sits squarely within the Benchmarking and Evaluation cluster, providing systematic assessment of how well existing methods actually generalize across time. Its emphasis on rigorous evaluation protocols aligns closely with Wild-time Benchmark[20] and Time Series OOD Survey[36], but distinguishes itself by critically examining whether reported gains reflect true temporal robustness or artifacts of evaluation design. This work addresses a fundamental question: are we measuring what we think we are measuring when we claim temporal generalization?

Related Works in Same Category

The following **3 sibling papers** share the same taxonomy leaf node with the original paper:

1. Wild-time: A benchmark of in-the-wild distribution shift over time

Authors: Yao, Huaxiu, Choi, Caroline, Cao, et al. (12 authors total) | **Year/Venue**: 2022 | **URL**: View paper

Abstract

Distribution shift occurs when the test distribution differs from the training distribution, and it can considerably degrade performance of machine learning models deployed in the real world. Temporal shifts -- distribution shifts arising from the passage of time -- often occur gradually and have the additional structure of timestamp metadata. By leveraging timestamp metadata, models can potentially learn from trends in past distribution shifts and extrapolate into the future. While recent works...

Relationship Analysis

Both papers belong to the Benchmarking and Evaluation category, focusing on curated datasets and evaluation protocols for assessing temporal robustness under distribution shifts. They overlap in their emphasis on evaluating model performance degradation over time using real-world temporal datasets (e.g., news, yearbook photos, satellite imagery). The key difference is that the original paper conducts a reality check on parameter interpolation and extrapolation methods for temporal generalization without future data access, while the candidate paper (Wild-Time) introduces a comprehensive benchmark with multiple datasets and evaluation strategies (Eval-Fix and Eval-Stream) to systematically compare various domain generalization, continual learning, and self-supervised learning approaches.

---

2. Understanding the Limits of Deep Tabular Methods with Temporal Shift

Authors: Ye, Han-Jia | **Year/Venue**: 2025 • International Conference on Machine Learning | **URL**: View paper

Abstract

Deep tabular models have demonstrated remarkable success on i.i.d. data, excelling in a variety of structured data tasks. However, their performance often deteriorates under temporal distribution shifts, where trends and periodic patterns are present in the evolving data distribution over time. In this paper, we explore the underlying reasons for this failure in capturing temporal dependencies. We begin by investigating the training protocol, revealing a key issue in how model selection performs...

Relationship Analysis

Both papers belong to the Benchmarking and Evaluation category, focusing on assessing model robustness under temporal distribution shifts. They overlap in examining how machine learning models degrade over time and evaluating methods to improve temporal generalization. However, the original paper conducts a broad empirical study across multiple domains (language modeling, summarization, classification) comparing parameter interpolation and extrapolation methods, while the candidate paper specifically focuses on deep tabular models, investigating training protocols and proposing temporal embeddings based on Fourier series to capture periodic patterns in tabular data.

---

3. Out-of-Distribution Generalization in Time Series: A Survey

Authors: Wu Xin, Teng Fei, Li, Xingwang, Zhang Ji, et al. (7 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

Time series frequently manifest distribution shifts, diverse latent features, and non-stationary learning dynamics, particularly in open and evolving environments. These characteristics pose significant challenges for out-of-distribution (OOD) generalization. While substantial progress has been made, a systematic synthesis of advancements remains lacking. To address this gap, we present the first comprehensive review of OOD generalization methodologies for time series, organized to delineate the...

Relationship Analysis

Both papers belong to the Benchmarking and Evaluation category, focusing on assessing temporal generalization under distribution shifts. The original paper provides a reality check by empirically evaluating parameter interpolation and extrapolation methods across diverse temporal tasks (language modeling, summarization, classification), finding that simple baselines often outperform complex approaches. The candidate paper is a comprehensive survey that systematically organizes the broader field of out-of-distribution generalization in time series, covering theoretical foundations, methodological categories (decoupling, invariant learning, adaptive mechanisms, large models), and evaluation protocols, rather than conducting empirical benchmarking experiments.

Contributions Analysis

Overall novelty summary. This paper contributes a systematic empirical evaluation of parameter interpolation and extrapolation methods for temporal generalization, examining whether models can generalize to future data using only past parameters. It resides in the 'Benchmarking and Evaluation' leaf under 'Theoretical Foundations and Benchmarking', alongside three sibling papers. This leaf represents a relatively sparse but critical research direction within the broader taxonomy of 50 papers across 18 leaf nodes, focusing specifically on evaluation protocols and benchmark design rather than novel adaptation algorithms or theoretical guarantees.

The taxonomy reveals that most research effort concentrates on developing adaptation methods (Test-Time Adaptation, Domain Adaptation branches contain 11 papers) and time series techniques (8 papers), while benchmarking work remains comparatively underexplored. The paper's neighboring leaves include 'Theoretical Analysis and Estimation' (3 papers on generalization bounds) and 'Model Selection and Assessment' (2 papers on validation strategies). Unlike these theoretical neighbors or the adaptation-focused branches, this work emphasizes empirical assessment of existing methods across diverse temporal tasks, bridging the gap between method development and rigorous evaluation of temporal robustness claims.

Among 27 candidates examined through limited semantic search, none clearly refute the paper's three main contributions. The first contribution (large-scale evaluation of parameter methods) examined 9 candidates with 0 refutable; the second (negative finding on method effectiveness) examined 8 with 0 refutable; the third (design principles identification) examined 10 with 0 refutable. This suggests that within the examined scope, the specific focus on parameter-space methods for temporal generalization and the systematic negative findings represent relatively unexplored territory, though the limited search scale means potentially relevant work may exist beyond these 27 candidates.

Based on this limited analysis of top-27 semantic matches, the work appears to occupy a distinct niche: systematic benchmarking of parameter-space approaches specifically for temporal shifts. The absence of refuting candidates within this scope, combined with the sparse population of the benchmarking leaf, suggests the contribution addresses an underserved evaluation need. However, the restricted search scope and the paper's focus on negative results warrant careful interpretation of its novelty claims relative to the broader literature.

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

Contribution 1: Large-scale empirical evaluation of parameter interpolation and extrapolation methods for temporal generalization

Description: The authors conduct a comprehensive empirical study comparing parameter interpolation methods (such as model merging and downscaling) and parameter extrapolation methods (such as Taylor-series approximation) across diverse temporal tasks and datasets, including language modeling, news summarization, classification tasks, and satellite imagery, using models ranging from 70M to 770M parameters under the strict constraint of no future data access.

This contribution was assessed against **9 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. CaT-GNN: Enhancing Credit Card Fraud Detection via Causal Temporal Graph Neural Networks

URL: View paper

Brief Assessment

Causal Temporal GNN[64] focuses on credit card fraud detection using causal graph neural networks, not on parameter interpolation/extrapolation methods for temporal generalization in machine learning models.

---

2. Physics-informed reduced order model with conditional neural fields

URL: View paper

Brief Assessment

Conditional Neural Fields[65] focuses on physics-informed reduced-order modeling for parametrized PDEs using neural ODEs and coordinate-based networks, not on parameter interpolation/extrapolation methods for temporal generalization in machine learning models across diverse tasks.

---

3. Un-mixing test-time normalization statistics: Combatting label temporal correlation

URL: View paper

Brief Assessment

Test-Time Normalization Statistics[63] focuses on test-time batch normalization adaptation for non-i.i.d. streaming data with label temporal correlation, not on parameter interpolation/extrapolation methods for temporal generalization across diverse tasks and datasets.

---

4. Continuous temporal domain generalization

URL: View paper

Brief Assessment

Continuous Temporal Generalization[61] focuses on continuous-time dynamics using Koopman operators for temporal domain generalization, not on empirical comparison of parameter interpolation/extrapolation methods across diverse tasks and datasets as in the original paper.

---

5. Graph neural processes for spatio-temporal extrapolation

URL: View paper

Brief Assessment

Graph Neural Processes[62] focuses on spatio-temporal extrapolation for sensor data in graphs using neural latent variable models, not on parameter interpolation/extrapolation methods for temporal generalization in machine learning models.

---

6. Training for the future: A simple gradient interpolation loss to generalize along time

URL: View paper

Brief Assessment

Gradient Interpolation Loss[67] focuses on a gradient interpolation training loss for temporal generalization, not on empirically evaluating parameter interpolation/extrapolation methods across diverse tasks and model scales (70M-770M parameters) as the original paper does.

---

7. A temporal-spatial interpolation and extrapolation method based on geographic Long Short-Term Memory neural network for PM2. 5

URL: View paper

Brief Assessment

Geographic LSTM Extrapolation[60] focuses on spatial-temporal interpolation/extrapolation for PM2.5 prediction using geographic data, not parameter-level interpolation/extrapolation methods for neural network temporal generalization across diverse ML tasks.

---

8. Drift-Resilient TabPFN: In-Context Learning Temporal Distribution Shifts on Tabular Data

URL: View paper

Brief Assessment

Drift-Resilient TabPFN[16] focuses on tabular data using in-context learning with prior-data fitted networks and structural causal models, not on parameter interpolation/extrapolation methods for neural network checkpoints across temporal tasks.

---

9. Temporal and geographic extrapolation of soil moisture using machine learning algorithms

URL: View paper

Brief Assessment

Soil Moisture Extrapolation[59] focuses on temporal and geographic extrapolation of soil moisture using machine learning algorithms for environmental prediction, not on parameter interpolation/extrapolation methods for neural network temporal generalization across diverse NLP and vision tasks.

Contribution 2: Empirical finding that parameter interpolation and extrapolation methods fail to consistently improve over the recent model baseline

Description: The authors demonstrate through extensive experiments that none of the evaluated temporal generalization methods reliably outperform the simple baseline of using the most recent model parameters, revealing the fundamental difficulty of predicting future model parameters from historical data alone without access to future distributions or strong assumptions about the data-generating process.

This contribution was assessed against **8 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Bam! just like that: Simple and efficient parameter upcycling for mixture of experts

URL: View paper

Brief Assessment

Parameter Upcycling MoE[53] focuses on initializing mixture-of-experts models from pre-trained dense models for improved performance, not on temporal generalization or parameter interpolation/extrapolation methods for handling distribution shifts over time.

2. How to Merge Multimodal Models Over Time?

URL: View paper

Brief Assessment

Multimodal Model Merging[52] focuses on temporal model merging for multimodal models (CLIP) in continual pretraining scenarios, not on the general temporal generalization problem across diverse tasks. The candidate studies merging strategies for initialization and deployment in continual learning, while the original paper examines whether parameter interpolation/extrapolation can predict future model parameters without access to future data across language modeling, summarization, and classification tasks.

3. : Cycle-Consistent Multi-Model Merging

URL: View paper

Brief Assessment

Cycle-Consistent Model Merging[54] focuses on merging models trained on the same task with different initializations through neuron permutation alignment, not on temporal generalization across evolving data distributions. The candidate does not address parameter interpolation/extrapolation for future time prediction.

4. A Systematic Study of Model Merging Techniques in Large Language Models

URL: View paper

Brief Assessment

Model Merging Techniques[55] focuses on merging fine-tuned LLM checkpoints for multi-task performance, not temporal generalization across time-evolving data distributions. The candidate evaluates model merging methods (task arithmetic, ties-merging, model stock) on static benchmarks, while the original paper addresses temporal distribution shifts where models must generalize to future time periods without access to future data.

5. KNOWLEDGE FUSION OF LARGE LANGUAGE MODELS VIA MODULAR SKILLPACKS

URL: View paper

Brief Assessment

Knowledge Fusion Skillpacks[58] focuses on cross-capability transfer between heterogeneous LLMs through knowledge distillation and modular skillpacks, not on temporal generalization or parameter interpolation/extrapolation methods for handling distribution shifts over time.

6. Curriculum Model Merging: Harmonizing Chemical LLMs for Enhanced Cross-Task Generalization

URL: View paper

Brief Assessment

Curriculum Model Merging[57] focuses on merging task-specific chemical LLMs through curriculum-based progressive merging, not on temporal generalization or parameter interpolation/extrapolation methods for future data prediction.

7. Validation approach for statistical extrapolation

URL: View paper

Brief Assessment

Statistical Extrapolation Validation[56] focuses on validation approaches for statistical extrapolation in shallow water scenarios with physical parameters, not on machine learning model parameter interpolation/extrapolation for temporal generalization tasks.

8. Machine Learning in Interpolation and Extrapolation for Nanophotonic Inverse Design

URL: View paper

Brief Assessment

Nanophotonic Inverse Design[51] focuses on interpolation and extrapolation of nanophotonic structure designs using neural networks for optical response prediction, not on temporal generalization of ML model parameters over time-evolving datasets.

Contribution 3: Identification of key design principles and challenges for temporal generalization

Description: The authors analyze the role of continual learning in maintaining parameter trajectory smoothness, the effect of parameter norm growth over time, and the challenges posed by non-identifiability and non-convexity in neural networks. They provide insights into hyperparameter selection without future data access and discuss fundamental theoretical constraints on temporal generalization.

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. Time matters: Empirical insights into the limits and challenges of temporal generalization in CSI-based Wi-Fi sensing

URL: View paper

Brief Assessment

CSI Temporal Generalization[71] focuses on Wi-Fi sensing applications (indoor positioning and human activity recognition) and how CSI patterns change over time due to physical signal propagation effects. The original paper addresses temporal generalization in machine learning models through parameter interpolation/extrapolation methods across diverse tasks (language modeling, image classification, etc.), which is a fundamentally different domain and approach.

---

2. Diversifying spatial-temporal perception for video domain generalization

URL: View paper

Brief Assessment

Video Domain Generalization[74] focuses on spatial-temporal diversity for video classification across domains, not on temporal generalization design principles for neural networks or parameter trajectory analysis over time.

---

3. Multi-view spatial-temporal graph convolutional networks with domain generalization for sleep stage classification

URL: View paper

Brief Assessment

Sleep Stage Generalization[70] focuses on spatial-temporal graph convolutions for sleep stage classification with domain generalization across subjects, not on temporal generalization of neural network parameters over time or design principles for maintaining model performance on future data distributions.

---

4. A comprehensive survey of deep learning for time series forecasting: architectural diversity and open challenges

URL: View paper

Brief Assessment

Time Series Forecasting Survey[69] focuses on architectural diversity in time series forecasting models and domain-specific challenges (distribution shift, causality, feature extraction), not on the specific design principles for temporal generalization in neural networks such as parameter trajectory smoothness, parameter norm growth, non-identifiability, or hyperparameter selection without future data access that the original paper investigates.

---

5. Characterizing the dynamics of mental representations: the temporal generalization method

URL: View paper

Brief Assessment

Temporal Generalization Method[76] focuses on characterizing temporal dynamics of mental representations in cognitive neuroscience, not on machine learning model parameter trajectories or neural network temporal generalization challenges.

---

6. Temporal Flexibility in Spiking Neural Networks: Towards Generalization Across Time Steps and Deployment Friendliness

URL: View paper

Brief Assessment

Temporal Flexibility Spiking Networks[77] addresses temporal flexibility in spiking neural networks across different time steps, not the broader temporal generalization challenges in machine learning models that the original paper investigates (e.g., distribution shifts over time, parameter trajectory smoothness, hyperparameter selection without future data).

---

7. Physics-informed neural networks for PDE problems: A comprehensive review

URL: View paper

Brief Assessment

Physics-Informed Neural Networks Review[73] focuses on PDE solvers using physics-informed neural networks for scientific computing problems, not on temporal generalization in machine learning models or design principles for handling distribution shifts over time.

---

8. A Residual Physics-Informed Neural Network Approach for Identifying Dynamic Parameters in Swing Equation-Based Power Systems

URL: View paper

Brief Assessment

Residual PINN Power Systems[75] focuses on parameter estimation in power systems using physics-informed neural networks with residual architectures. It does not address temporal generalization in machine learning models or the design principles for maintaining performance under distribution shifts over time.

---

9. 4d spatio-temporal convnets: Minkowski convolutional neural networks

URL: View paper

Brief Assessment

Minkowski Convolutional Networks[68] focuses on 4D spatio-temporal convolutions for 3D video perception (e.g., depth images, lidar scans), not on temporal generalization principles for neural networks under distribution shifts. The candidate addresses architectural design for processing spatial-temporal data, while the original contribution analyzes theoretical constraints like parameter trajectory smoothness and hyperparameter selection for temporal generalization.

---

10. Learning Latent Spaces for Domain Generalization in Time Series Forecasting

URL: View paper

Brief Assessment

Latent Spaces Domain Generalization[72] focuses on domain generalization in time series forecasting through latent factor learning, not on analyzing design principles for temporal generalization in neural networks or continual learning dynamics.

Appendix: Text Similarity Detection

Textual similarity detection checked 31 papers and found 1 similarity segment(s) across 1 paper(s).

The following **1 paper(s)** were detected to have high textual similarity with the original paper. These may represent different versions of the same work, duplicate submissions, or papers with substantial textual overlap. Readers are advised to verify these relationships independently.

1. Wild-time: A benchmark of in-the-wild distribution shift over time

Detected in: Core Task (sibling)

⚠ **Note**: This paper shows substantial textual similarity with the original paper. It may be a different version, a duplicate submission, or contain significant overlapping content. Please review carefully to determine the nature of the relationship.

References

- [0] Temporal Generalization: A Reality Check View paper
- [1] CERA: A Framework for Improved Generalization of Machine Learning Models to Changed Climates View paper
- [2] Reversible instance normalization for accurate time-series forecasting against distribution shift View paper
- [3] Diversify: A general framework for time series out-of-distribution detection and generalization View paper
- [4] Adaptive State Estimation and Continual Learning under Data Distribution Shift View paper
- [5] Generalizable Recommender System During Temporal Popularity Distribution Shifts View paper
- [6] Bayestta: Continual-temporal test-time adaptation for vision-language models via gaussian discriminant analysis View paper
- [7] Dynamic graph neural networks under spatio-temporal distribution shift View paper
- [8] Online test-time adaptation for better generalization of interatomic potentials to out-of-distribution data View paper
- [9] Calibration of time-series forecasting: Detecting and adapting context-driven distribution shift View paper
- [10] Adaptable: Test-time adaptation for tabular data via shift-aware uncertainty calibrator and label distribution handler View paper
- [11] Improving the Generalization of Segmentation Foundation Model under Distribution Shift via Weakly Supervised Adaptation View paper
- [12] Temporal Restoration and Spatial Rewiring for Source-Free Multivariate Time Series Domain Adaptation View paper
- [13] A Multi-modal Architecture with Spatio-Temporal-Text Adaptation for Video-based Traffic Accident Anticipation View paper
- [14] Adaptive Conformal Inference Under Distribution Shift View paper
- [15] Simple temporal adaptation to changing label sets: Hashtag prediction via dense KNN View paper
- [16] Drift-Resilient TabPFN: In-Context Learning Temporal Distribution Shifts on Tabular Data View paper
- [17] Predicting Practically? Domain Generalization for Predictive Analytics in Real-world Environments View paper
- [18] Improving Generalization of Dynamic Graph Learning via Environment Prompt View paper
- [19] Adaptive Personalized Federated Learning for Non-IID Data with Continual Distribution Shift View paper
- [20] Wild-time: A benchmark of in-the-wild distribution shift over time View paper
- [21] Learning pattern-specific experts for time series forecasting under patch-level distribution shift View paper
- [22] Towards Streaming Land Use Classification of Images with Temporal Distribution Shifts View paper
- [23] Temporal Distribution Shift in Real-World Pharmaceutical Data: Implications for Uncertainty Quantification in QSAR Models View paper
- [24] Model Assessment and Selection under Temporal Distribution Shift View paper
- [25] EHR foundation models improve robustness in the presence of temporal distribution shift View paper
- [26] Universal Domain Adaptation for Robust Handling of Distributional Shifts in NLP View paper
- [27] Graph learning under distribution shifts: A comprehensive survey on domain adaptation, out-of-distribution, and continual learning View paper
- [28] Evaluation of domain generalization and adaptation on improving model robustness to temporal dataset shift in clinical medicine View paper
- [29] Robust Multivariate Time Series Forecasting Against Intraseries and Interseries Transitional Shift View paper
- [30] Understanding the Limits of Deep Tabular Methods with Temporal Shift View paper
- [31] Learning to Align: Addressing Character Frequency Distribution Shifts in Handwritten Text Recognition View paper
- [32] A survey of deep graph learning under distribution shifts: from graph out-of-distribution generalization to adaptation View paper
- [33] Test-time adaptation to distribution shift by confidence maximization and input transformation View paper
- [34] Adaptive temporal transformer method for short-term wind power forecasting considering shift in time series distribution View paper
- [35] Robust Machine Learning: Detection, Evaluation and Adaptation Under Distribution Shift View paper
- [36] Out-of-Distribution Generalization in Time Series: A Survey View paper
- [37] Distributional drift adaptation with temporal conditional variational autoencoder for multivariate time series forecasting View paper
- [38] Dual prototype evolving for test-time generalization of vision-language models View paper
- [39] Efficient Non-stationary Online Learning by Wavelets with Applications to Online Distribution Shift Adaptation View paper
- [40] Temporal Evaluation ofÂ Uncertainty Quantification Under Distribution Shift View paper
- [41] Asp: Learn a universal neural solver! View paper
- [42] Handling new class in online label shift View paper
- [43] Adaptive streaming architectures: Integrating change-data-capture with online machine learning for real-time enterprise decisioning View paper
- [44] Open Spatio-Temporal Foundation Models for Traffic Prediction View paper
- [45] CFTResNet: A novel cross-domain diagnosis framework guided by interpretability for cardiovascular diseases. View paper
- [46] Source-free domain adaptation with temporal imputation for time series data View paper
- [47] Monitoring of Laser Powder Bed Fusion process by bridging dissimilar process maps using deep learning-based domain adaptation on acoustic emissions View paper
- [48] Assessing and Mitigating Medical Knowledge Drift and Conflicts in Large Language Models View paper
- [49] Adaptive Estimation and Learning under Temporal Distribution Shift View paper
- [50] Deep Learning-Based Financial Fraud Detection with Temporal and Feature-Level Adaptation View paper
- [51] Machine Learning in Interpolation and Extrapolation for Nanophotonic Inverse Design View paper
- [52] How to Merge Multimodal Models Over Time? View paper
- [53] Bam! just like that: Simple and efficient parameter upcycling for mixture of experts View paper
- [54] : Cycle-Consistent Multi-Model Merging View paper
- [55] A Systematic Study of Model Merging Techniques in Large Language Models View paper

- [56] Validation approach for statistical extrapolation View paper
- [57] Curriculum Model Merging: Harmonizing Chemical LLMs for Enhanced Cross-Task Generalization View paper
- [58] KNOWLEDGE FUSION OF LARGE LANGUAGE MODELS VIA MODULAR SKILLPACKS View paper
- [59] Temporal and geographic extrapolation of soil moisture using machine learning algorithms View paper
- [60] A temporal-spatial interpolation and extrapolation method based on geographic Long Short-Term Memory neural network for PM2. 5 View paper
- [61] Continuous temporal domain generalization View paper
- [62] Graph neural processes for spatio-temporal extrapolation View paper
- [63] Un-mixing test-time normalization statistics: Combatting label temporal correlation View paper
- [64] CaT-GNN: Enhancing Credit Card Fraud Detection via Causal Temporal Graph Neural Networks View paper
- [65] Physics-informed reduced order model with conditional neural fields View paper
- [66] Online video deblurring via dynamic temporal blending network View paper
- [67] Training for the future: A simple gradient interpolation loss to generalize along time View paper
- [68] 4d spatio-temporal convnets: Minkowski convolutional neural networks View paper
- [69] A comprehensive survey of deep learning for time series forecasting: architectural diversity and open challenges View paper
- [70] Multi-view spatial-temporal graph convolutional networks with domain generalization for sleep stage classification View paper
- [71] Time matters: Empirical insights into the limits and challenges of temporal generalization in CSI-based Wi-Fi sensing View paper
- [72] Learning Latent Spaces for Domain Generalization in Time Series Forecasting View paper
- [73] Physics-informed neural networks for PDE problems: A comprehensive review View paper
- [74] Diversifying spatial-temporal perception for video domain generalization View paper
- [75] A Residual Physics-Informed Neural Network Approach for Identifying Dynamic Parameters in Swing Equation-Based Power Systems View paper
- [76] Characterizing the dynamics of mental representations: the temporal generalization method View paper
- [77] Temporal Flexibility in Spiking Neural Networks: Towards Generalization Across Time Steps and Deployment Friendliness View paper