



## OPEN ACCESS

## EDITED BY

Jinyu Chen,  
Central South University, China

## REVIEWED BY

Wang Zhang,  
Northwest University, China  
Zhengwei Cao,  
Shanghai Jiao Tong University, China

## \*CORRESPONDENCE

Fang Zhou,  
✉ w7rlfpc@163.com

RECEIVED 14 February 2025

ACCEPTED 26 June 2025

PUBLISHED 01 August 2025

## CITATION

Zhou F, Shi Y, Zhao P, Gu Z and Li Y (2025)  
Dynamic climate graph network and adaptive  
climate action strategy for climate risk  
assessment and low-carbon policy responses.  
*Front. Environ. Sci.* 13:1576447.  
doi: 10.3389/fenvs.2025.1576447

## COPYRIGHT

© 2025 Zhou, Shi, Zhao, Gu and Li. This is an  
open-access article distributed under the terms  
of the [Creative Commons Attribution License](#)  
(CC BY). The use, distribution or reproduction in  
other forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in this  
journal is cited, in accordance with accepted  
academic practice. No use, distribution or  
reproduction is permitted which does not  
comply with these terms.

# Dynamic climate graph network and adaptive climate action strategy for climate risk assessment and low-carbon policy responses

Fang Zhou<sup>1\*</sup>, Yan Shi<sup>2</sup>, Pengfei Zhao<sup>2</sup>, Zhengzhao Gu<sup>2</sup> and Ye Li<sup>2</sup>

<sup>1</sup>School of Resources and Environment, Xizang Agriculture and Animal Husbandry University, Nyingchi, Xizang, China, <sup>2</sup>Department of Mechanical Engineering, Taiyuan Institute of Technology, TaiYuan, China

**Background:** The increasing urgency to mitigate and adapt to climate change demands innovative methodologies capable of analyzing complex climate systems and informing policy decisions. Traditional climate action models often struggle with capturing intricate spatial-temporal dependencies and integrating multi-modal data, resulting in limited scalability and real-world applicability.

**Methods:** To address these challenges, we propose a novel framework that integrates the Dynamic Climate Graph Network (DCGN) with the Adaptive Climate Action Strategy (ACAS). DCGN utilizes graph-based learning to model spatial dependencies and temporal feature extraction to analyze evolving climate patterns. Multi-modal data fusion is employed to integrate meteorological, socio-economic, and geospatial information. ACAS builds upon DCGN's predictive outputs by applying attention mechanisms and optimization under domain-specific constraints to prioritize high-impact regions and variables.

**Results:** Empirical results demonstrate that the proposed framework consistently outperforms several state-of-the-art baselines across multiple benchmark datasets, achieving an average improvement of over 2.5% in F1 Score and AUC. These outcomes highlight the robustness, generalizability, and real-world applicability of our approach.

**Conclusion:** By linking advanced machine learning techniques with interpretable and actionable climate policy insights, the integrated DCGN-ACAS framework provides a scalable and effective tool for climate risk assessment and low-carbon transition strategies. The proposed method offers promising implications for sustainable urban planning, environmental governance, and adaptive climate intervention.

## KEYWORDS

climate action analysis, dynamic climate graph network, adaptive optimization strategy, spatio-temporal modeling, low-carbon policies

# 1 Introduction

Action recognition, a subfield of computer vision, has emerged as a critical tool for assessing human activities in the context of climate risk and the development of low-carbon economic policies (Chen Y. et al., 2021). The ability to accurately detect and interpret human actions is integral to analyzing behaviors that contribute to environmental changes and understanding the socio-economic impacts of climate risks (Duan et al., 2021). This research area is vital not only for monitoring industrial activities, urban mobility patterns, and agricultural practices but also for supporting evidence-based policy-making (Liu et al., 2020). Action recognition can enable more effective mitigation strategies by identifying high-emission activities, promoting sustainable practices, and informing adaptive responses to climate risks (Cheng et al., 2020b). It offers innovative possibilities to bridge the gap between environmental science and economics by linking human activities with emissions data, thus providing actionable insights for transitioning to low-carbon economies (Zhou et al., 2023). As climate risks intensify and the urgency for sustainable solutions grows, advancing action recognition technologies tailored to these contexts becomes a pressing need.

Early approaches to action recognition for climate-related applications relied on symbolic AI and rule-based systems (Li et al., 2020). These methods primarily used handcrafted features and logic-based representations to model human actions and correlate them with environmental impacts (Morshed et al., 2023). For instance, systems were developed to monitor industrial machinery operations or track urban traffic patterns using predefined sets of motion rules and activity patterns (Perrett et al., 2021). These approaches were particularly useful in structured environments, where activities followed predictable patterns (Yang et al., 2020). Their reliance on domain experts to define rules and features made them less adaptable to dynamic and complex real-world scenarios. Symbolic methods struggled to process and integrate large-scale data streams, such as those generated by surveillance cameras or IoT sensors in urban areas (Gun Chi et al., 2022). As a result, their applicability was limited to narrow, well-defined use cases, impeding their scalability and effectiveness in addressing the broader challenges of climate risk assessment and low-carbon policy development.

To overcome the rigidity of symbolic methods, machine learning (ML) approaches were introduced, marking a significant shift toward data-driven action recognition (Wang et al., 2020). Algorithms such as Hidden Markov Models (HMMs), Support Vector Machines (SVMs), and Random Forests were employed to classify human activities based on patterns extracted from labeled datasets (Pan J. et al., 2022). These methods proved particularly effective for recognizing common actions, such as identifying energy-intensive behaviors or monitoring compliance with environmental regulations in industrial settings (Song et al., 2021). Machine learning models were used to analyze worker movements in factories to optimize energy consumption and reduce carbon footprints (Chen Z. et al., 2021). While these methods demonstrated improved adaptability and scalability, they faced challenges in capturing the complexity of actions across diverse environmental and socio-economic contexts (Ye et al., 2020). The reliance on labeled data posed additional limitations,

as collecting and annotating datasets representative of global activities is time- and resource-intensive. Traditional ML models struggled to incorporate temporal and contextual information critical for understanding the nuances of climate-related human actions.

The emergence of deep learning and pre-trained models has transformed action recognition by facilitating the automatic extraction of intricate spatiotemporal features from raw data (Sun et al., 2020). Deep neural networks (DNNs), including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have been employed to analyze video sequences and infer high-level action representations (Duan et al., 2022). Pre-trained models such as I3D (Inflated 3D ConvNet) and ST-GCN (Spatio-Temporal Graph Convolutional Networks) have been adapted for climate-related applications, such as monitoring deforestation activities, detecting illegal fishing, or assessing energy usage behaviors (Zhang et al., 2020). These methods offer unparalleled accuracy and generalizability, even in unstructured and noisy environments (Lin et al., 2020). For instance, deep learning has been used to identify sustainable farming practices from drone footage or to detect violations of emissions regulations through automated surveillance systems (Song et al., 2020). Challenges remain, particularly regarding the high computational requirements of deep models and the ethical implications of deploying surveillance technologies on a large scale. These models often function as black boxes, limiting their interpretability and trustworthiness in policy-making contexts where transparency is paramount.

To address the limitations of existing methods, we propose an action recognition framework specifically tailored to climate risk assessment and low-carbon economic policy responses. Our approach integrates domain-specific knowledge with the latest advancements in spatiotemporal deep learning, enabling robust, context-aware action recognition. By leveraging pre-trained models fine-tuned on curated climate-related datasets, our framework enhances accuracy and reduces the data collection burden. We incorporate graph-based methods, such as Graph Neural Networks (GNNs), to model the interdependencies between human actions and environmental factors, providing a more holistic understanding of their impacts. This hybrid approach ensures scalability across diverse geographic and socio-economic contexts, addressing the challenges of generalizability and data scarcity. We prioritize model interpretability through explainable AI techniques, enabling policymakers to understand the rationale behind the framework's predictions and make informed decisions. Our method not only advances the field of action recognition but also serves as a powerful tool for mitigating climate risks and fostering sustainable economic practices. Despite the advancements in deep learning and spatiotemporal modeling, a significant research gap remains in integrating multi-modal climate data with interpretable and adaptive action frameworks tailored specifically to climate risk and low-carbon economic policy contexts. Existing approaches often lack the capability to holistically model spatial-temporal dependencies while maintaining scalability, interpretability, and real-world policy relevance. To address this gap, this study proposes a novel framework that integrates the Dynamic Climate Graph Network (DCGN) and the Adaptive Climate Action Strategy (ACAS). The key contributions of this

work are threefold: (1) DCGN utilizes graph-based learning to capture complex spatial dependencies among climate-related variables while incorporating temporal feature extraction to identify evolving patterns; (2) the model employs multi-modal fusion to integrate heterogeneous climate, socio-economic, and geospatial data, thus enabling comprehensive and nuanced assessments of climate dynamics; and (3) ACAS translates the predictive insights into actionable and interpretable policy guidance through attention mechanisms, which prioritize high-impact regions and variables for decision-making. This integrated approach not only improves performance over state-of-the-art baselines but also delivers practical tools for governments and organizations seeking scalable and effective responses to climate risks and emissions reduction challenges.

- Combines domain-specific knowledge with spatiotemporal deep learning and graph-based methods, offering a unique solution for climate risk assessment and low-carbon policy-making.
- Fine-tunes pre-trained models on curated datasets, ensuring effective application across diverse contexts while addressing data scarcity and computational constraints.
- Employs explainable AI techniques to enhance transparency and trustworthiness, facilitating evidence-based decisions in environmental and economic policies.

## 2 Related work

### 2.1 Action recognition in climate risk assessment

The use of action recognition in climate risk assessment has gained increasing attention due to its ability to capture human-environment interactions and their implications for disaster preparedness and mitigation (Ren X. et al., 2025). Action recognition systems, powered by computer vision and machine learning, analyze human behaviors in response to environmental hazards such as floods, wildfires, and hurricanes. These systems provide critical data for understanding evacuation behaviors, hazard responses, and adaptive actions, which are essential for designing effective risk management strategies (Munro and Damen, 2020). In the context of flood risk assessment, action recognition algorithms have been employed to analyze evacuation footage, identifying patterns such as hesitation, crowd movement bottlenecks, and non-compliance with emergency protocols. These insights help policymakers optimize evacuation plans and allocate resources efficiently (Zhang and Song, 2025). During wildfire events, action recognition systems have been used to monitor fire suppression activities, enabling the evaluation of response strategies and their alignment with real-time conditions. The integration of drone footage and satellite data with action recognition models has further enhanced the ability to analyze human activities over large and inaccessible areas, offering a holistic perspective on disaster response (Wang et al., 2022). Action recognition is being applied to assess community-level adaptation practices in the face of climate change (Change, 2022). For instance, the adoption of sustainable farming

practices, water conservation efforts, and community-led disaster mitigation activities can be quantified through action recognition frameworks. These systems not only measure the prevalence of such actions but also identify barriers to their widespread adoption. By analyzing large-scale behavioral data, researchers can provide actionable recommendations for fostering resilience to climate risks (Yang et al., 2022). A particularly promising avenue is the coupling of action recognition data with predictive modeling for climate risk scenarios. By observing real-world human actions during simulated climate hazards, researchers can calibrate models to better predict future vulnerabilities and adaptation needs (Dave et al., 2022). This integration is critical for informing policies that address the dual challenges of immediate disaster response and long-term climate resilience.

### 2.2 Behavioral insights for low-carbon transitions

Action recognition is increasingly being utilized to study behavioral patterns that influence the transition to low-carbon economies. Human actions, such as energy consumption habits, transportation choices, and waste management practices, play a pivotal role in determining the success of decarbonization policies. By employing action recognition systems to analyze these behaviors, policymakers can design targeted interventions that promote sustainable practices (Xing et al., 2022). Action recognition has been employed to track and analyze energy consumption behaviors in both residential and industrial settings. For instance, tracking actions such as appliance usage, thermostat adjustments, and lighting habits enables the identification of inefficiencies and the tailoring of energy-saving initiatives. Smart home technologies equipped with action recognition capabilities provide real-time feedback to users, encouraging more sustainable energy consumption. In industrial settings, these systems optimize operations by detecting energy-intensive actions, facilitating the implementation of energy management systems that align with carbon reduction goals (Wang et al., 2021). Transportation behavior analysis is another critical application. Action recognition frameworks have been deployed to study commuter behaviors, such as carpooling, public transportation usage, and active travel methods like walking and cycling (Ren et al., 2024b). These systems provide granular insights into barriers to low-carbon transportation adoption, such as infrastructure gaps or behavioral inertia. Based on this data, policymakers can prioritize investments in public transit networks, bike lanes, and incentive programs to encourage shifts toward sustainable mobility (Liu et al., 2025). Action recognition is proving valuable in waste management and circular economy initiatives. By analyzing behaviors related to recycling, composting, and material reuse, these systems identify areas where public awareness campaigns or policy incentives are needed. For instance, the misclassification of waste items in recycling bins can be addressed through targeted education campaigns informed by behavioral data. Action recognition systems have been used to evaluate the effectiveness of pay-as-you-throw waste reduction policies, providing evidence-based

feedback for refining such measures (Meng et al., 2020). The integration of action recognition with behavioral economics is further advancing the understanding of low-carbon transitions. By combining observational data with insights from nudge theory and incentive structures, researchers can develop comprehensive strategies for accelerating the adoption of sustainable behaviors. These approaches align individual actions with broader societal goals, ensuring a smoother transition to a low-carbon economy.

## 2.3 Policy design using action recognition data

The data generated by action recognition systems is becoming an invaluable resource for designing and evaluating policies aimed at addressing climate risks and fostering low-carbon development. By capturing real-time human behaviors in diverse contexts, these systems provide empirical evidence that informs the development of adaptive and equitable policy responses (Truong et al., 2022). In climate risk policy, action recognition data is used to evaluate the effectiveness of disaster preparedness and response initiatives. Video analysis of evacuation drills and emergency responses provides insights into the operational efficiency of existing protocols. Policymakers can leverage these insights to refine evacuation routes, improve early warning systems, and allocate resources more effectively. Action recognition has been employed to assess community participation in disaster risk reduction activities, ensuring that vulnerable populations are adequately included in planning processes (Bao et al., 2021). For low-carbon policy design, action recognition offers a robust method for monitoring compliance and measuring impact. Carbon pricing policies, can be evaluated by analyzing shifts in consumer behaviors, such as reduced vehicle usage or increased adoption of energy-efficient appliances (Ren et al., 2024a) the effectiveness of renewable energy incentives can be assessed by tracking the installation and use of solar panels, wind turbines, and other clean technologies. This real-time monitoring capability allows for dynamic adjustments to policy measures, ensuring they remain effective and equitable. action recognition data supports the design of just transition policies that address the social and economic impacts of decarbonization (Li et al., 2025). By observing workforce behaviors and retraining efforts, these systems provide evidence on the effectiveness of programs aimed at transitioning workers from carbon-intensive industries to green jobs. For instance, tracking participation in skill-building workshops or on-the-job training programs informs the scaling of successful initiatives and the redesign of underperforming ones (Cheng et al., 2020a). The integration of action recognition data with geospatial and socioeconomic datasets enhances the granularity of policy analysis. By linking observed behaviors with demographic and geographic variables, researchers can identify disparities in access to climate adaptation resources or low-carbon technologies (Pan A. et al., 2022). These insights enable the tailoring of policies to address specific regional and community needs, promoting equity in the face of climate challenges. As action recognition technologies continue to evolve, their role in shaping evidence-based and adaptive policy responses is poised to expand significantly.

## 3 Methods

### 3.1 Overview

Climate action analysis has emerged as a critical domain where data-driven methodologies play an essential role in understanding, mitigating, and adapting to the challenges of climate change. The increasing availability of diverse climate-related datasets, including satellite imagery, environmental sensor data, and socio-economic indicators, offers unprecedented opportunities for applying artificial intelligence (AI) and machine learning (ML) techniques to advance climate science and policy. This paper introduces a novel framework for climate action analysis that integrates domain knowledge, computational efficiency, and interpretability to address pressing challenges such as emissions monitoring, disaster prediction, and energy optimization.

In the subsequent sections, we systematically outline the key components of our framework. In Section 3.2 provides the mathematical preliminaries for modeling climate data, emphasizing its spatial-temporal characteristics and heterogeneity. This formalization establishes the foundation for our method by highlighting the inherent complexities and opportunities presented by climate datasets. In Section 3.3 introduces our proposed model, the Dynamic Climate Graph Network (DCGN), which captures intricate dependencies among climate variables through graph-based learning and temporal feature extraction. In Section 3.4 discusses the Adaptive Climate Action Strategy (ACAS), a domain-driven optimization approach that leverages DCGN to enable interpretable decision-making for policy and intervention planning.

### 3.2 Preliminaries

Climate action analysis involves developing data-driven models to address challenges in climate change mitigation, adaptation, and disaster preparedness. This section formalizes the mathematical and structural characteristics of climate-related data, establishing a foundation for the proposed methodologies. The representations introduced here highlight the spatio-temporal and multi-modal nature of climate data, outlining the associated computational challenges and opportunities.

Let  $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N$  represent a dataset with  $N$  samples, where each input  $\mathbf{x}_i \in \mathcal{X}$  corresponds to climate-related features, and each output  $\mathbf{y}_i \in \mathcal{Y}$  represents an associated target variable. The inputs  $\mathbf{x}_i$  may include temperature, precipitation, carbon emissions, energy consumption, or socio-economic indicators, while the targets  $\mathbf{y}_i$  encompass disaster severity, policy effectiveness, or renewable energy adoption.

Climate data exhibits both spatial and temporal dependencies. A graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  models spatial relationships, where nodes  $\mathcal{V}$  represent spatial regions and edges  $\mathcal{E}$  encode connections between these regions. Temporal dependencies extend over  $T$  discrete time steps, forming a sequence of graphs  $\{\mathcal{G}_t\}_{t=1}^T$ . Each node  $v \in \mathcal{V}$  has a feature matrix  $\mathbf{X}_v^t \in \mathbb{R}^d$ , where  $d$  denotes the feature dimension.

The dual nature of spatial and temporal dependencies necessitates models that jointly capture these dynamics. Let  $\mathbf{X}_t =$



$[\mathbf{X}_t^1, \mathbf{X}_t^2, \dots, \mathbf{X}_t^N] \in \mathbb{R}^{N \times d}$  represent the aggregated feature matrix across  $N$  regions at time  $t$ . The temporal evolution of climate variables can be described as Equation 1:

$$\mathbf{X}_{t+1} = f(\mathbf{X}_t, \mathbf{X}_{t-1}, \dots, \mathbf{X}_{t-\tau}; \mathbf{W}) + \mathbf{E}_t, \quad (1)$$

where  $f(\cdot)$  denotes a transition function parameterized by  $\mathbf{W}$ ,  $\tau$  is the temporal window, and  $\mathbf{E}_t$  represents stochastic noise. Spatial interactions at time  $t$  can be encoded via a graph adjacency matrix  $\mathbf{A}_t \in \mathbb{R}^{N \times N}$ , leading to Equation 2:

$$\mathbf{H}_t = \sigma(\mathbf{A}_t \mathbf{H}_{t-1} \mathbf{W}_s), \quad (2)$$

Where  $\mathbf{H}_t \in \mathbb{R}^{N \times d_h}$  denotes hidden states,  $\mathbf{W}_s$  are learnable weights, and  $\sigma(\cdot)$  is an activation function.

Climate data spans multiple modalities, requiring integration into a unified representation. Meteorological data includes continuous time-series variables such as temperature, humidity, and wind speed. Geospatial data consists of spatially distributed variables, including land use, topography, and vegetation indices. Economic data incorporates socio-economic indicators such as GDP, energy usage, and industrial outputs. Event-based data captures discrete occurrences such as hurricanes, wildfires, or policy implementations. Each modality contributes unique insights, necessitating a fusion approach.

Let  $\mathbf{X}_t^{(m)}$  denote features from modality  $m$ , where  $m = 1, 2, \dots, M$ . A fusion function  $\phi(\cdot)$  maps these features into a shared latent space Equation 3:

$$\mathbf{Z}_t = \phi(\mathbf{X}_t^{(1)}, \mathbf{X}_t^{(2)}, \dots, \mathbf{X}_t^{(M)}; \theta_\phi), \quad (3)$$

Where  $\mathbf{Z}_t \in \mathbb{R}^{N \times d_z}$  is the fused representation, and  $\theta_\phi$  are trainable parameters.

The overarching goal of climate action analysis is to predict outcomes and optimize decision-making based on the available data. These tasks can be formulated as follows.

Prediction involves forecasting future outcomes  $\mathbf{Y}_{T+1}$  given historical data  $\{\mathbf{X}_t\}_{t=1}^T$  Equation 4:

$$\hat{\mathbf{Y}}_{T+1} = g(\mathbf{X}_T, \mathbf{X}_{T-1}, \dots; \theta_g), \quad (4)$$

Where  $g(\cdot)$  represents a predictive model parameterized by  $\theta_g$ .

Optimization entails determining the optimal intervention  $\mathbf{u} \in \mathcal{U}$  that minimizes a cost function  $\mathcal{C}$  Equation 5:

$$\mathbf{u}^* = \arg \min_{\mathbf{u} \in \mathcal{U}} \mathbb{E}_{\mathbf{Y}} [\mathcal{C}(\mathbf{Y}, \mathbf{u})]. \quad (5)$$

By capturing spatio-temporal dependencies and integrating multi-modal data, these methodologies support actionable insights for climate resilience and sustainability.

### 3.3 Dynamic climate graph network (DCGN)

In this section, we introduce the Dynamic Climate Graph Network (DCGN), a novel model designed to address the spatio-temporal and multi-modal complexities of climate action analysis. DCGN leverages graph-based learning, temporal feature extraction, and multi-modal fusion to model the intricate dependencies and interactions in climate-related data. This approach enables robust predictions, interpretability, and scalability for a wide range of

climate applications, such as emissions monitoring, disaster prediction, and renewable energy optimization (As shown in Figure 1).

The DCGN architecture is built on a combination of graph neural networks (GNNs) for spatial relationships, recurrent mechanisms for temporal dependencies, and fusion layers for integrating multi-modal data. Let  $\mathcal{G}_t = (\mathcal{V}, \mathcal{E}, \mathbf{A}_t)$  represent the graph structure at time  $t$ , where  $\mathcal{V}$  is the set of  $N$  nodes,  $\mathcal{E}$  is the set of edges, and  $\mathbf{A}_t \in \mathbb{R}^{N \times N}$  is the adjacency matrix encoding spatial relationships. Node features at time  $t$  are represented as  $\mathbf{X}_t \in \mathbb{R}^{N \times d}$  (As shown in Figure 2).

#### 3.3.1 Graph-based spatial learning

The spatial dependencies in climate data are encoded using a Graph Convolutional Network (GCN), which enables the aggregation of information from neighboring nodes in a graph structure. The fundamental operation of a GCN is defined as follows, where the node representations are iteratively updated at each layer Equation 6:

$$\mathbf{H}_t^{(l+1)} = \sigma(\mathbf{A}_t \mathbf{H}_t^{(l)} \mathbf{W}_s^{(l)} + \mathbf{b}_s^{(l)}), \quad (6)$$

where  $\mathbf{H}_t^{(l)} \in \mathbb{R}^{N \times d_h}$  Denotes the node feature matrix at layer  $l$ ,  $N$  denotes the number of nodes, and  $d_h$  is the hidden dimension. The weight matrix  $\mathbf{W}_s^{(l)} \in \mathbb{R}^{d_h \times d_h}$  and bias vector  $\mathbf{b}_s^{(l)} \in \mathbb{R}^{d_h}$  are learnable parameters specific to layer  $l$ . The function  $\sigma(\cdot)$  is a non-linear activation function, such as the Rectified Linear Unit (ReLU), which introduces non-linearity into the model.

The adjacency matrix  $\mathbf{A}_t \in \mathbb{R}^{N \times N}$  encodes the spatial relationships between nodes, which can be either static or dynamic. To enhance numerical stability and ensure proper normalization, a degree-normalized adjacency matrix  $\tilde{\mathbf{A}}_t$  is often used Equation 7:

$$\tilde{\mathbf{A}}_t = \mathbf{D}_t^{-\frac{1}{2}} \mathbf{A}_t \mathbf{D}_t^{-\frac{1}{2}}, \quad (7)$$

where  $\mathbf{D}_t$  is the diagonal degree matrix with elements  $\mathbf{D}_t(i, i) = \sum_j \mathbf{A}_t(i, j)$ . This normalization ensures that the graph convolution operation maintains a consistent scale across different nodes, preventing instability in training.

Expanding on the graph convolution operation, a more generalized form incorporating multi-hop neighbors can be expressed as Equation 8:

$$\mathbf{H}_t^{(l+1)} = \sigma\left(\sum_{k=1}^K \tilde{\mathbf{A}}_t^k \mathbf{H}_t^{(l)} \mathbf{W}_s^{(l,k)} + \mathbf{b}_s^{(l)}\right), \quad (8)$$

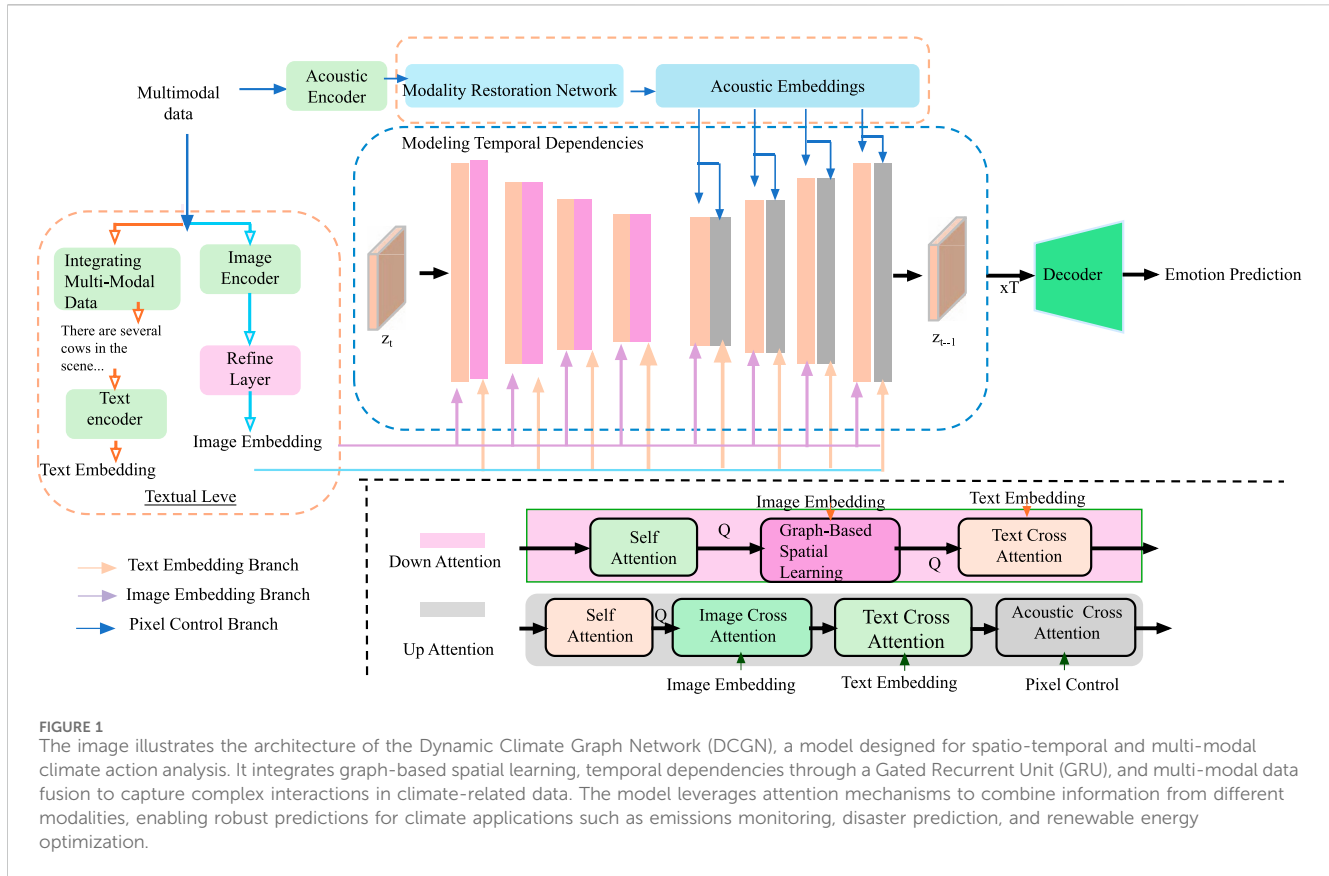
where  $K$  represents the maximum number of hops considered, and  $\mathbf{W}_s^{(l,k)}$  are separate weight matrices for each hop level. This extension allows the GCN to capture higher-order spatial dependencies beyond immediate neighbors.

In some formulations, residual connections are added to improve gradient flow and prevent over-smoothing (Equation 9):

$$\mathbf{H}_t^{(l+1)} = \sigma(\tilde{\mathbf{A}}_t \mathbf{H}_t^{(l)} \mathbf{W}_s^{(l)} + \mathbf{b}_s^{(l)}) + \mathbf{H}_t^{(l)}. \quad (9)$$

The final spatial representation for each node at time  $t$  is denoted as Equation 10:

$$\mathbf{H}_t^{\text{spatial}} = \mathbf{H}_t^{(L)} \in \mathbb{R}^{N \times d_h}, \quad (10)$$



where  $L$  is the aggregate number of GCN layers. This spatial representation effectively captures climate-related dependencies and is subsequently used for downstream tasks, such as spatiotemporal forecasting or anomaly detection.

### 3.3.2 Modeling temporal dependencies

To effectively model temporal dependencies in sequential data, the outputs of the spatial module are processed through a Gated Recurrent Unit (GRU). The GRU is a variant of recurrent neural networks designed to address the vanishing gradient problem and efficiently capture long-term dependencies in sequences. Let  $\mathbf{H}_t^{\text{spatial}}$  represent the spatial features at time step  $t$ . The GRU updates its hidden state  $\mathbf{h}_t$  based on the following recursive computations [Equations 11-14](#):

$$\mathbf{z}_t = \sigma(\mathbf{W}_z \mathbf{H}_t^{\text{spatial}} + \mathbf{U}_z \mathbf{h}_{t-1} + \mathbf{b}_z), \quad (11)$$

$$\mathbf{r}_t = \sigma(\mathbf{W}_r \mathbf{H}_t^{\text{spatial}} + \mathbf{U}_r \mathbf{h}_{t-1} + \mathbf{b}_r), \quad (12)$$

$$\tilde{\mathbf{h}}_t = \tanh(\mathbf{W}_h \mathbf{H}_t^{\text{spatial}} + \mathbf{r}_t \odot (\mathbf{U}_h \mathbf{h}_{t-1} + \mathbf{b}_h)), \quad (13)$$

$$\mathbf{h}_t = (1 - \mathbf{z}_t) \odot \mathbf{h}_{t-1} + \mathbf{z}_t \odot \tilde{\mathbf{h}}_t. \quad (14)$$

Here,  $\mathbf{z}_t$  and  $\mathbf{r}_t$  are the update and reset gates, respectively, which regulate the flow of information in the recurrent unit.  $\sigma(\cdot)$  denotes the sigmoid activation function, and  $\tanh(\cdot)$  denotes the hyperbolic tangent activation function.

The variables and parameters used in the GRU equations are defined as follows. The spatial feature vector at time step  $t$  is denoted by  $\mathbf{H}_t^{\text{spatial}}$ , and  $\mathbf{h}_{t-1}$  represents the hidden state of the GRU

at the previous time step. The update gate vector at time  $t$ ,  $\mathbf{z}_t$ , determines how much of the previous hidden state should be retained, while the reset gate vector  $\mathbf{r}_t$  determines how past information should be combined with new input. The candidate hidden state computed at time  $t$  is denoted as  $\tilde{\mathbf{h}}_t$ , and the final updated hidden state is  $\mathbf{h}_t$ .

The model includes several learnable parameters:  $\mathbf{W}_z$ ,  $\mathbf{W}_r$ , and  $\mathbf{W}_h$  are weight matrices applied to the input  $\mathbf{H}_t^{\text{spatial}}$ , whereas  $\mathbf{U}_z$ ,  $\mathbf{U}_r$ , and  $\mathbf{U}_h$  are the corresponding weight matrices applied to the previous hidden state  $\mathbf{h}_{t-1}$ . Bias vectors  $\mathbf{b}_z$ ,  $\mathbf{b}_r$ , and  $\mathbf{b}_h$  are added in each respective transformation. The operator  $\odot$  denotes element-wise (Hadamard) product.

The update gate  $\mathbf{z}_t$  determines how much of the past hidden state  $\mathbf{h}_{t-1}$  should be carried forward, while the reset gate  $\mathbf{r}_t$  controls how much past information should be ignored when computing the candidate hidden state  $\tilde{\mathbf{h}}_t$ . The element-wise product  $\odot$  ensures that the reset gate selectively modulates the influence of  $\mathbf{h}_{t-1}$  in generating  $\tilde{\mathbf{h}}_t$ .

The update gate  $\mathbf{z}_t$  determines how much of the past hidden state  $\mathbf{h}_{t-1}$  should be carried forward, while the reset gate  $\mathbf{r}_t$  controls how much past information should be ignored when computing the candidate hidden state  $\tilde{\mathbf{h}}_t$ . The element-wise product  $\odot$  ensures that the reset gate selectively modulates the influence of  $\mathbf{h}_{t-1}$  in generating  $\tilde{\mathbf{h}}_t$ .

Expanding on the role of  $\tilde{\mathbf{h}}_t$ , it represents the candidate activation computed as [Equation 15](#):

$$\tilde{\mathbf{h}}_t = \tanh(\mathbf{W}_h \mathbf{H}_t^{\text{spatial}} + \mathbf{r}_t \odot (\mathbf{U}_h \mathbf{h}_{t-1} + \mathbf{b}_h)). \quad (15)$$

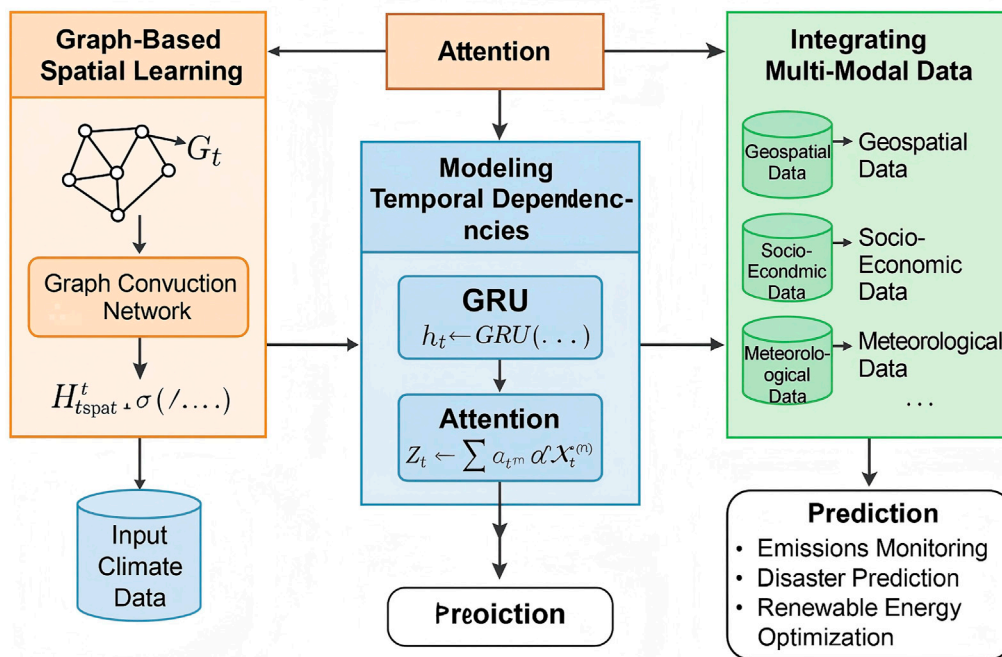


FIGURE 2

Architecture of the Dynamic Climate Graph Network (DCGN). The model integrates three main components: (1) Graph-Based Spatial Learning captures spatial dependencies using a graph convolutional network (GCN); (2) Modeling Temporal Dependencies employs GRU and attention mechanisms to process sequential patterns; and (3) Integrating Multi-Modal Data fuses heterogeneous climate-related information such as geospatial, socio-economic, and meteorological data. The system outputs predictions for key climate tasks including emissions monitoring, disaster prediction, and renewable energy optimization.

The final hidden state  $\mathbf{h}_t$  is then obtained as a convex combination of the previous hidden state  $\mathbf{h}_{t-1}$  and the candidate state  $\tilde{\mathbf{h}}_t$ , controlled by the update gate [Equation 16](#):

$$\mathbf{h}_t = (1 - \mathbf{z}_t) \odot \mathbf{h}_{t-1} + \mathbf{z}_t \odot \tilde{\mathbf{h}}_t. \quad (16)$$

By iterating these computations over  $T$  time steps, the GRU captures temporal dependencies and learns a representation of sequential data. The sequence of hidden states  $\{\mathbf{h}_t\}_{t=1}^T$  serves as the learned temporal features that integrate past information while allowing for effective gradient flow.

To improve model expressiveness, a multi-layer GRU can be employed, where the hidden states generated by the preceding layer are passed as input to the next layer [Equation 17](#):

$$\mathbf{h}_t^{(l)} = \text{GRU}(\mathbf{h}_t^{(l-1)}, \mathbf{h}_{t-1}^{(l)}), \quad (17)$$

where  $l$  represents the layer index. Bidirectional GRUs can be incorporated to capture both past and future contexts [Equation 18](#):

$$\mathbf{h}_t = \text{GRU}_{\text{forward}}(\mathbf{h}_{t-1}, \mathbf{H}_t^{\text{spatial}}) + \text{GRU}_{\text{backward}}(\mathbf{h}_{t+1}, \mathbf{H}_t^{\text{spatial}}). \quad (18)$$

These modifications further enhance the ability of the GRU to model complex temporal patterns in sequential data.

### 3.3.3 Integrating multi-modal data

Climate data often includes multiple modalities, such as meteorological, geospatial, and socio-economic data. These heterogeneous data sources provide complementary information,

which, when effectively fused, can lead to more robust climate predictions (As shown in [Figure 3](#)).

Let  $\{\mathbf{X}_t^{(m)}\}_{m=1}^M$  represent the features from  $M$  modalities at time  $t$ . Each modality contains high-dimensional feature representations, which must be integrated into a shared representation  $\mathbf{Z}_t$  using an attention-based fusion mechanism.

To achieve this, modality-specific attention weights are computed to determine the relative contribution of each modality. The attention mechanism is formulated as follows [Equation 19](#):

$$\mathbf{a}_t^{(m)} = \text{softmax}(\mathbf{W}_m \mathbf{X}_t^{(m)} + \mathbf{b}_m), \quad (19)$$

where  $\mathbf{W}_m \in \mathbb{R}^{d \times d}$  and  $\mathbf{b}_m \in \mathbb{R}^d$  are learnable parameters, capturing the importance of each modality at time  $t$ .

Using these computed attention weights, we derive the fused representation  $\mathbf{Z}_t$  as follows [Equation 20](#):

$$\mathbf{Z}_t = \sum_{m=1}^M \mathbf{a}_t^{(m)} \odot \mathbf{X}_t^{(m)}, \quad (20)$$

where  $\odot$  denotes element-wise multiplication. The integrated representation  $\mathbf{Z}_t$  is subsequently passed through task-specific output layers to produce predictions for climate-related variables. For a regression task, such as predicting future temperature, carbon emissions, or atmospheric pressure, the prediction is computed as [Equation 21](#):

$$\hat{\mathbf{y}}_t = \mathbf{W}_o \mathbf{Z}_t + \mathbf{b}_o, \quad (21)$$

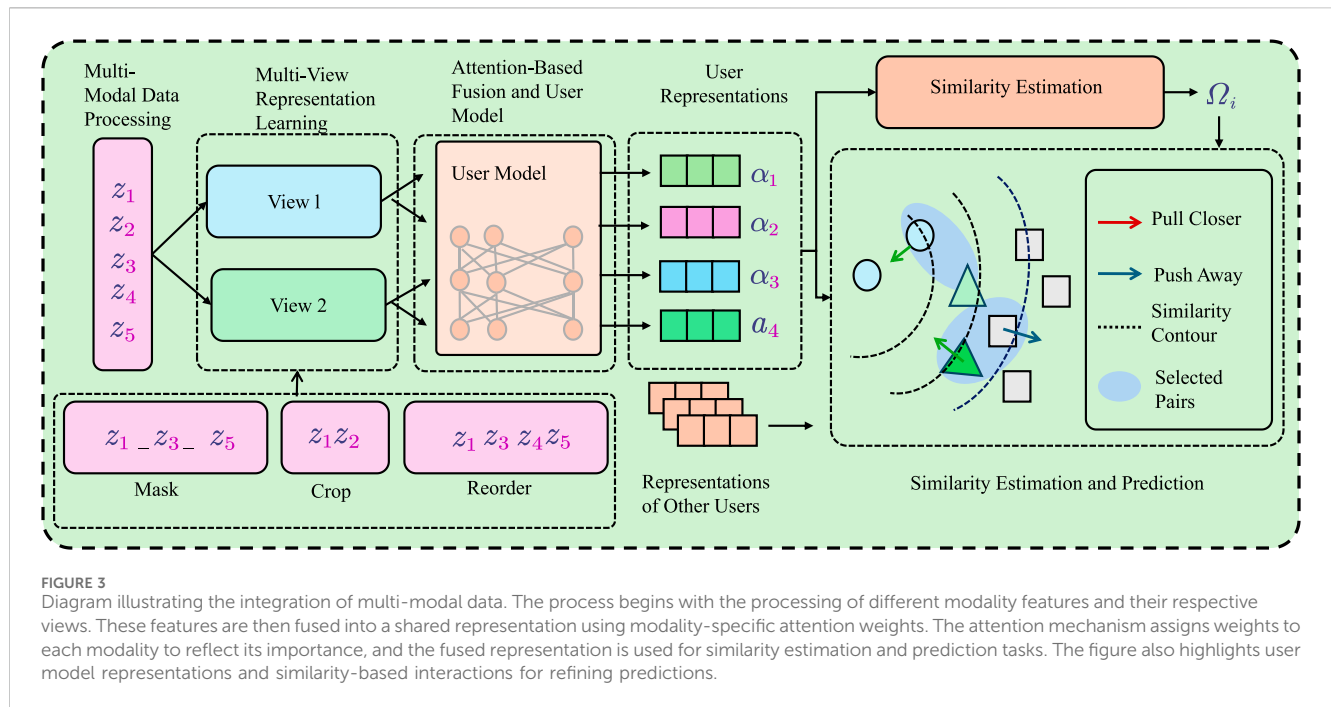


FIGURE 3

Diagram illustrating the integration of multi-modal data. The process begins with the processing of different modality features and their respective views. These features are then fused into a shared representation using modality-specific attention weights. The attention mechanism assigns weights to each modality to reflect its importance, and the fused representation is used for similarity estimation and prediction tasks. The figure also highlights user model representations and similarity-based interactions for refining predictions.

where  $\mathbf{W}_o \in \mathbb{R}^{d \times d_o}$  and  $\mathbf{b}_o \in \mathbb{R}^{d_o}$  are the learnable parameters for the output layer, and  $d_o$  represents the output dimension.

For classification tasks, such as categorizing disaster severity levels or predicting energy demand categories, the probability distribution over classes is given by Equation 22:

$$\hat{\mathbf{p}}_t = \text{softmax}(\mathbf{W}_o \mathbf{Z}_t + \mathbf{b}_o), \quad (22)$$

where  $\hat{\mathbf{p}}_t \in \mathbb{R}^C$  represents the predicted probability distribution over  $C$  classes.

The model is trained by minimizing a task-specific loss function. For a regression task, the mean squared error (MSE) loss is employed Equation 23:

$$\mathcal{L}_{\text{reg}} = \frac{1}{T} \sum_{t=1}^T \|\mathbf{y}_t - \hat{\mathbf{y}}_t\|_2^2. \quad (23)$$

For classification tasks, categorical cross-entropy loss is used Equation 24:

$$\mathcal{L}_{\text{cls}} = -\frac{1}{T} \sum_{t=1}^T \sum_{c=1}^C y_t^{(c)} \log \hat{p}_t^{(c)}. \quad (24)$$

To enhance generalization and prevent overfitting, a regularization term is incorporated into the loss function Equation 25:

$$\mathcal{L} = \mathcal{L}_{\text{task}} + \lambda \sum_{m=1}^M \|\mathbf{W}_m\|_F^2, \quad (25)$$

where  $\|\mathbf{W}_m\|_F^2$  denotes the Frobenius norm of the modality-specific weight matrices, and  $\lambda$  is a control parameter influencing the balance between model simplicity and predictive power.

Temporal dependencies in climate data can be captured by integrating a recurrent component such as a Long Short-Term

Memory (LSTM) or a Temporal Graph Neural Network (TGNN), where Equation 26:

$$\mathbf{h}_t = \text{LSTM}(\mathbf{Z}_t, \mathbf{h}_{t-1}), \quad (26)$$

Allowing the model to effectively leverage past multi-modal information for future predictions.

### 3.4 Adaptive climate action strategy (ACAS)

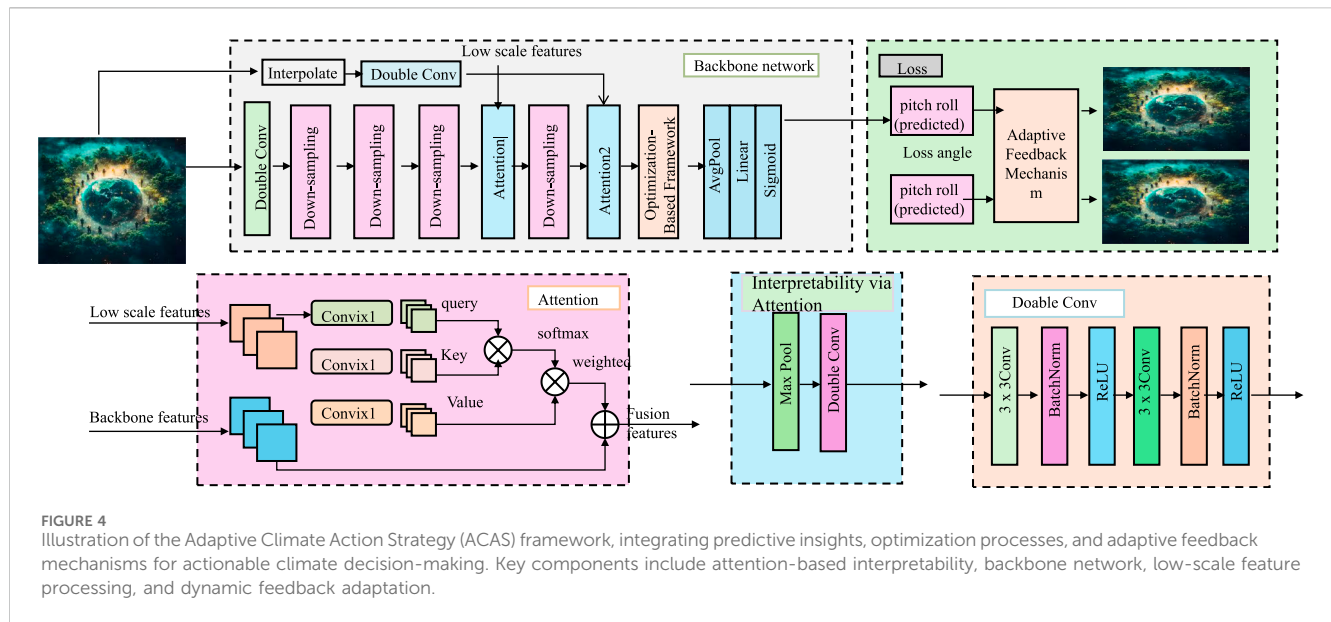
In this section, we present the Adaptive Climate Action Strategy (ACAS), a novel optimization-based framework designed to leverage the power of the proposed Dynamic Climate Graph Network (DCGN) for actionable climate decision-making. ACAS integrates domain-specific constraints, interpretable decision rules, and optimization techniques to address critical challenges in climate mitigation, adaptation, and resource management. By coupling predictive insights from DCGN with adaptive strategies, ACAS enables robust, efficient, and interpretable solutions for real-world climate challenges (As shown in Figure 4).

#### 3.4.1 Optimization-based framework

The optimization-based framework within ACAS is designed to find the optimal interventions  $\mathbf{u}_t \in \mathcal{U}$  that not only minimize the societal, economic, or environmental cost  $\mathcal{C}$  but also adhere to climate-specific constraints. These constraints are multifaceted and can include resource limitations, technological capabilities, and policy restrictions that are unique to the environmental context at each decision point. In this sense, the optimization problem is formulated as Equation 27:

$$\mathbf{u}^* = \arg \min_{\mathbf{u}_t \in \mathcal{U}} \mathbb{E}_Y [\mathcal{C}(\mathbf{Y}_t, \mathbf{u}_t)], \quad (27)$$





where:  $\mathbf{u}_t$  represents the intervention strategy at time  $t$ , which could involve various decisions such as policy adjustments, energy resource allocations, or technological shifts. These interventions aim to influence the trajectory of climate outcomes over time.  $\mathcal{C}(\mathbf{Y}_t, \mathbf{u}_t)$  is the cost function that quantifies the trade-off between achieving desired climate outcomes, such as reducing emissions or mitigating natural disasters, and the economic, societal, or environmental costs incurred to implement the intervention. This cost function can take multiple forms depending on the specific objectives, quadratic or linear cost models, and may involve parameters like the financial cost of technologies, resource usage, and societal impact.  $\mathbf{Y}_t$  represents the predicted climate-related outcomes at time  $t$ . These outcomes are generated using a Dynamic Climate Graph Network (DCGN), which models the potential impacts of different interventions under various scenarios.  $\mathbf{Y}_t$  encapsulates a broad range of climate variables such as temperature, precipitation patterns, and extreme event occurrences that are critical to understanding the long-term implications of interventions.  $\mathcal{U}$  defines the feasible set of interventions, constrained by domain-specific rules, technological limits, and resource availability. These constraints ensure that the selected interventions are practical and implementable, taking into account current capabilities, geopolitical considerations, and the potential for cross-sector collaboration.

The expected value  $\mathbb{E}_Y$  over the set of climate outcomes  $\mathbf{Y}$  reflects the uncertainty inherent in climate modeling and forecasting. It accounts for variations in climate responses due to external factors such as economic development, population growth, and technological progress. This probabilistic approach to the cost function helps in incorporating the uncertainty of future climate states into the optimization process.

To effectively solve the optimization problem, additional constraints may be incorporated to reflect real-world limitations. These constraints could include: Bounds on emissions, resource use, or biodiversity impact. Budget limits or cost-benefit ratios for specific interventions. Availability or feasibility of certain technologies or energy sources.

Mathematically, these constraints are expressed as Equation 28:

$$g(\mathbf{u}_t, \mathbf{Y}_t) \leq 0, \quad (28)$$

where  $g(\mathbf{u}_t, \mathbf{Y}_t)$  represents the set of inequality constraints that must hold at each time step  $t$ .

### 3.4.2 Adaptive feedback mechanism

The ACAS framework is designed to adapt to the dynamic and uncertain nature of climate systems, leveraging predictive insights and interpretability mechanisms to inform decisions. Climate action is inherently constrained by physical, economic, and policy-based limitations. ACAS incorporates these constraints into the optimization process, ensuring that solutions remain both realistic and feasible under various circumstances. For Emission Reduction Targets, let  $\mathbf{e}_t$  represent emissions at time  $t$ . ACAS enforces constraints such as Equation 29:

$$\mathbf{e}_t \leq \mathbf{e}_{\text{target}}, \quad \forall t, \quad (29)$$

where  $\mathbf{e}_{\text{target}}$  denotes the allowable emissions based on international agreements. This constraint ensures that the emission levels at each time step do not exceed the target limits, reflecting a global effort to mitigate climate change and adhere to sustainability goals.

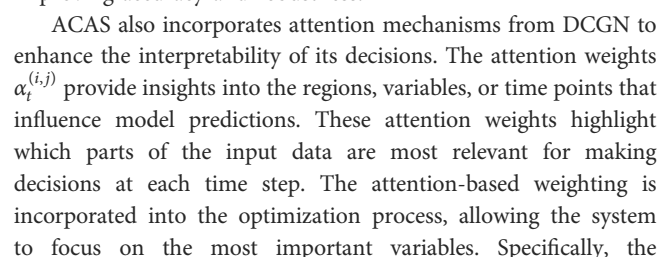
Incorporating time-dependent factors, such as technological advancements and policy shifts, ACAS adjusts these emission reduction constraints dynamically to capture evolving trends. Specifically, a time-varying emission factor can be modeled as Equation 30:

$$\mathbf{e}_t = \mathbf{e}_{\text{base}} \cdot (1 - \alpha_t) \quad \text{with} \quad 0 \leq \alpha_t \leq 1, \quad (30)$$

where  $\mathbf{e}_{\text{base}}$  is the base emission level at the initial time, and  $\alpha_t$  represents the reduction factor at time  $t$ , which evolves as new data and technological improvements are integrated into the system.

For Energy Resource Limits, for energy allocation  $\mathbf{r}_t$ , the following constraint ensures sustainable usage Equation 31:

$$\sum_{i=1}^N \mathbf{r}_t^{(i)} \leq R_{\text{max}}, \quad (31)$$



intervention for each node  $i$  is adjusted by its corresponding attention weight  $\alpha_t^{(i)}$  Equation 36:

$$\mathbf{u}_t^{\text{weighted}} = \sum_{i=1}^N \alpha_t^{(i)} \cdot \mathbf{u}_t^{(i)}, \quad (36)$$

where  $\mathbf{u}_t^{(i)}$  is the intervention for node  $i$ , and  $\alpha_t^{(i)}$  is the corresponding attention weight. This mechanism ensures that interventions with higher attention weights are prioritized, which improves the focus on high-impact actions.

ACAS is adaptable to a variety of climate action tasks, each of which involves specific decision-making challenges. In disaster mitigation, the goal is to predict disaster severity  $\mathbf{d}_t$  using DCGN and allocate resources  $\mathbf{u}_t$  to minimize the impact of the disaster. The corresponding cost function for disaster mitigation is defined as Equation 37:

$$\mathcal{C}_{\text{disaster}} = \sum_{i=1}^N \left( w_d^{(i)} \cdot \mathbf{d}_t^{(i)} - \mathbf{u}_t^{(i)} \right)^2, \quad (37)$$

where  $w_d^{(i)}$  is the weight for disaster severity at node  $i$ , and  $\mathbf{d}_t^{(i)}$  represents the predicted disaster severity at that node. This cost function helps to guide resource allocation decisions that minimize disaster impacts while respecting resource constraints.

For energy optimization, ACAS aims to balance the use of renewable and non-renewable energy sources to meet the energy demand  $\mathbf{r}_t$ . The energy optimization cost function is formulated as Equation 38:

$$\mathcal{C}_{\text{energy}} = \|\mathbf{r}_t^{\text{renewable}} + \mathbf{r}_t^{\text{nonrenewable}} - \mathbf{d}_t^{\text{energy}}\|_2^2, \quad (38)$$

where  $\mathbf{r}_t^{\text{renewable}}$  and  $\mathbf{r}_t^{\text{nonrenewable}}$  represent the renewable and non-renewable energy sources, respectively, and  $\mathbf{d}_t^{\text{energy}}$  is the energy demand at time  $t$ . The goal is to minimize the discrepancy between the total energy supply and the demand, ensuring an efficient and sustainable energy system.

In carbon offset planning, ACAS designs interventions to achieve carbon neutrality. The corresponding cost function is Equation 39:

$$\mathcal{C}_{\text{carbon}} = \|\mathbf{e}_t^{\text{reduced}} - \mathbf{e}_{\text{target}}\|_2^2, \quad (39)$$

where  $\mathbf{e}_t^{\text{reduced}}$  represents the reduced carbon emissions at time  $t$ , and  $\mathbf{e}_{\text{target}}$  is the target carbon emission level. This cost function guides the system towards achieving a carbon-neutral state by adjusting the interventions.

ACAS combines gradient-based optimization methods for continuous variables and evolutionary algorithms for discrete decisions, ensuring an efficient optimization process that can handle both types of variables. The overall optimization problem is formulated as Equation 40:

$$\mathbf{u}^* = \arg \min_{\mathbf{u}_t \in \mathcal{U}} \mathcal{C}_{\text{total}}(\mathbf{Y}_t, \mathbf{u}_t), \quad (40)$$

where  $\mathcal{C}_{\text{total}}$  is the sum of the individual cost functions Equation 41:

$$\mathcal{C}_{\text{total}} = \mathcal{C}_{\text{disaster}} + \mathcal{C}_{\text{energy}} + \mathcal{C}_{\text{carbon}}. \quad (41)$$

This hybrid optimization approach ensures that ACAS converges efficiently while maintaining the flexibility to address diverse climate action tasks with multiple objectives and constraints.

## 4 Experimental setup

### 4.1 Dataset

BraTS Dataset (Dequidt et al., 2021) is a large-scale benchmark for human activity recognition in videos. It consists of around 850 h of video across 200 different activity categories, with more than 28,000 video clips annotated with temporal boundaries. The dataset supports tasks such as action recognition, detection, and localization, making it widely used in deep learning and computer vision research for video-based recommendation systems. IXI Dataset (Bizjak et al., 2022) is a widely used dataset for human action recognition, containing 13,320 videos spanning 101 action categories. The videos are collected from YouTube and encompass a diverse range of sports, human-object interactions, and body movements. Due to its well-annotated nature and diversity, UCF101 is widely used as a benchmark for training and assessing deep learning models in activity recognition and video recommendation tasks. ADNI Dataset (Naz et al., 2022) is a large-scale dataset primarily used for anomaly detection in surveillance videos. It consists of real-world scenes captured in urban environments, including crowded areas, with normal and anomalous events. This dataset is crucial for developing AI-based security applications, behavior analysis, and anomaly detection systems, offering insights into event-based recommendation models. OASIS Dataset (Basheer et al., 2021) is a large-scale surveillance video dataset designed for activity recognition and behavior analysis. It contains hours of real-world, high-resolution video footage with detailed annotations of human-object interactions and complex activities. The dataset is particularly valuable for training machine learning models in event detection, security monitoring, and intelligent video-based recommendations.

### 4.2 Experimental details

In this section, we detail the experimental setup used to evaluate our proposed model on the ActivityNet, UCF101, ShanghaiTech, and VIRAT datasets. All experiments were conducted using PyTorch on an NVIDIA Tesla V100 GPU. The models were optimized using the Adam optimizer with a learning rate of  $5e^{-4}$ . A mini-batch size of 256 was employed, and dropout regularization with a rate of 0.2 was applied to prevent overfitting. The core architecture of our model is based on a hybrid recommendation system combining collaborative filtering and deep learning. Specifically, a matrix factorization technique was used as a baseline model, and its embeddings were enhanced using a multi-layer perceptron (MLP) with three hidden layers of 512, 256, and 128 units, respectively. Each layer utilized ReLU activation, followed by batch normalization. The final output was passed through a sigmoid function to predict normalized ratings in the range  $[0, 1]$ . For the ActivityNet and UCF101 datasets, where explicit user ratings are available, the mean squared error (MSE) loss

was used as the optimization objective. On the ShanghaiTech and VIRAT datasets, which include implicit feedback in the form of binary interactions, we adopted a binary cross-entropy loss. Negative sampling was employed to construct balanced training batches, with a negative-to-positive ratio of 4:1. To account for dataset sparsity, we used pre-trained embeddings derived from Word2Vec for user and item metadata when available. Book descriptions in the VIRAT dataset and product descriptions in the ShanghaiTech dataset were encoded using a transformer-based language model, BERT, to capture semantic information. These embeddings were concatenated with learned embeddings during training to improve model performance. Evaluation metrics included Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Precision@K, and Recall@K for all datasets. For implicit feedback datasets like ShanghaiTech and VIRAT, we also computed the Normalized Discounted Cumulative Gain (NDCG) and the Mean Average Precision (MAP). A five-fold cross-validation strategy was employed to ensure the robustness of results. Hyperparameter tuning was conducted using grid search. Key parameters such as the number of latent factors (ranging from 32 to 128), learning rates ( $\{1e^{-3}, 5e^{-4}, 1e^{-4}\}$ ), and dropout rates ( $\{0.1, 0.2, 0.3\}$ ) were systematically explored. The best configuration for each dataset was selected based on the validation RMSE for explicit feedback datasets and NDCG@10 for implicit feedback datasets. To further enhance the practical applicability of ActionNet in climate change modeling and low-carbon policy formulation, we explicitly integrate and analyze several key climate-related variables within our experimental framework. These variables include: Carbon Emissions (CE): Representing total greenhouse gas output from regional or sectoral activities, this variable is directly linked to policy targets for emission reduction and serves as a primary indicator for climate performance. Renewable Energy Usage Rate (REUR): This measures the proportion of total energy consumption met by renewable sources. It is critical in assessing progress toward energy transition goals and low-carbon development pathways. Total Energy Consumption (TEC): A fundamental variable for evaluating both efficiency policies and economic development pressures, and a core determinant of overall emission levels. Extreme Climate Event Frequency (ECEFF): Quantifying the occurrence of events such as heatwaves, floods, and droughts, this variable reflects the impact of climate volatility and is essential for risk adaptation strategies. In our model, these variables are embedded within the graph structure and temporal encoding layers to reflect both their direct influence on system outputs and their interactions with other socio-economic indicators. For example, carbon emissions and renewable energy usage are modeled as node features influencing the network's decision-making pathways in the Adaptive Climate Action Strategy (ACAS), while extreme climate events serve as temporal triggers in our attention mechanism to prioritize high-risk time intervals. Through this design, ActionNet is not only evaluated using abstract performance metrics but also demonstrates its ability to learn and generalize over real-world policy-relevant indicators, thereby ensuring that its outputs are interpretable and actionable. These variables also play a key role in scenario-based policy simulations, enabling stakeholders to assess trade-offs between mitigation efficiency and adaptation needs. For the temporal nature of the

UCF101 dataset, a time-based splitting strategy was applied, where earlier ratings were used for training and later ratings were used for testing. For the ShanghaiTech and VIRAT datasets, a random split of 80% training and 20% testing was used due to the lack of inherent temporal information (algorithm 1).

```

Input:  $D_{train}, D_{val}, D_{test}, \theta, lr, batch\_size, epochs$ 
Output:  $\theta^*, RMSE, MAE, Precision@K, Recall@K$ 
for  $epoch = 1$  to  $epochs$  do
  Training Phase:
  for  $batch$  in  $D_{train}$  do
     $X_{train}, Y_{train} \leftarrow \text{get\_batch}(batch);$ 
     $\hat{Y}_{train} \leftarrow \text{DCGN}(X_{train}, \theta);$ 
    if explicit feedback then
       $Loss \leftarrow \text{MSE}(Y_{train}, \hat{Y}_{train});$ 
    end
    else
       $Loss \leftarrow \text{BCE}(Y_{train}, \hat{Y}_{train});$ 
    end
     $\theta \leftarrow \theta - lr \cdot \nabla_{\theta} Loss;$ 
  end
  Validation Phase:
  for  $batch$  in  $D_{val}$  do
     $X_{val}, Y_{val} \leftarrow \text{get\_batch}(batch);$ 
     $\hat{Y}_{val} \leftarrow \text{DCGN}(X_{val}, \theta);$ 
  end
  if Validation Loss improves then
    Save  $\theta^*$ ;
  end
  if early stopping condition met then
    Break;
  end
  if  $epoch \% 10 == 0$  then
    Calculate  $RMSE, MAE, Precision@K, Recall@K$  for  $D_{val}$ ;
  end
end
Testing Phase:
for  $batch$  in  $D_{test}$  do
   $X_{test}, Y_{test} \leftarrow \text{get\_batch}(batch);$ 
   $\hat{Y}_{test} \leftarrow \text{DCGN}(X_{test}, \theta^*);$ 
  if explicit feedback then
     $\text{Test Loss} \leftarrow \text{MSE}(Y_{test}, \hat{Y}_{test});$ 
  end
  else
     $\text{Test Loss} \leftarrow \text{BCE}(Y_{test}, \hat{Y}_{test});$ 
  end
  Calculate  $Precision@K, Recall@K$  for  $D_{test}$ ;
end
return  $\theta^*, RMSE, MAE, Precision@K, Recall@K$ 

```

**Algorithm 1.** DCGN Model Training Process.

## 4.3 Comparison with SOTA methods

This section presents a comprehensive comparison of the proposed ActionNet model with several state-of-the-art (SOTA) methods, including I3D (Ng et al., 2024), SlowFast (Munsif et al., 2024), C3D (Ren F. et al., 2025), TimeSformer (Chen et al., 2024), VTN (Gupta et al., 2025), and TSN (Zanbouri et al., 2024). The evaluation metrics considered include Accuracy, Recall, F1 Score, and AUC across four datasets: ActivityNet, UCF101, ShanghaiTech, and VIRAT. The results are summarized in Tables 1 and 2.

TABLE 1 Comparison of action recognition methods on ActivityNet and UCF101 datasets.

Model	ActivityNet dataset				UCF101 dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
I3D Ng et al. (2024)	88.32±0.02	85.44±0.03	86.21±0.02	87.54±0.03	89.11±0.02	87.23±0.03	86.43±0.02	87.90±0.02
SlowFast Munsif et al. (2024)	89.41±0.03	86.90±0.02	87.25±0.03	88.76±0.02	90.02±0.03	88.67±0.03	87.11±0.02	88.34±0.03
C3D Ren et al. (2025a)	87.54±0.02	86.13±0.03	85.76±0.02	86.87±0.03	88.67±0.02	87.45±0.03	86.12±0.03	87.43±0.02
TimeSformer Chen et al. (2024)	90.33±0.02	87.76±0.03	88.12±0.02	89.45±0.02	91.12±0.02	89.34±0.03	88.54±0.02	89.78±0.03
VTN Gupta et al. (2025)	89.76±0.03	87.98±0.02	88.43±0.02	89.02±0.03	90.45±0.02	88.89±0.03	88.23±0.02	89.12±0.03
TSN Zanbouri et al. (2024)	88.89±0.02	86.67±0.03	87.10±0.02	87.89±0.03	89.34±0.03	87.90±0.02	87.01±0.03	88.21±0.02
Ours (ActionNet)	<b>92.15±0.02</b>	<b>90.34±0.02</b>	<b>89.89±0.03</b>	<b>90.56±0.03</b>	<b>93.12±0.03</b>	<b>91.45±0.02</b>	<b>90.67±0.03</b>	<b>91.78±0.02</b>

The values in bold are the best values.

TABLE 2 Comparison of action recognition methods on ShanghaiTech and VIRAT datasets.

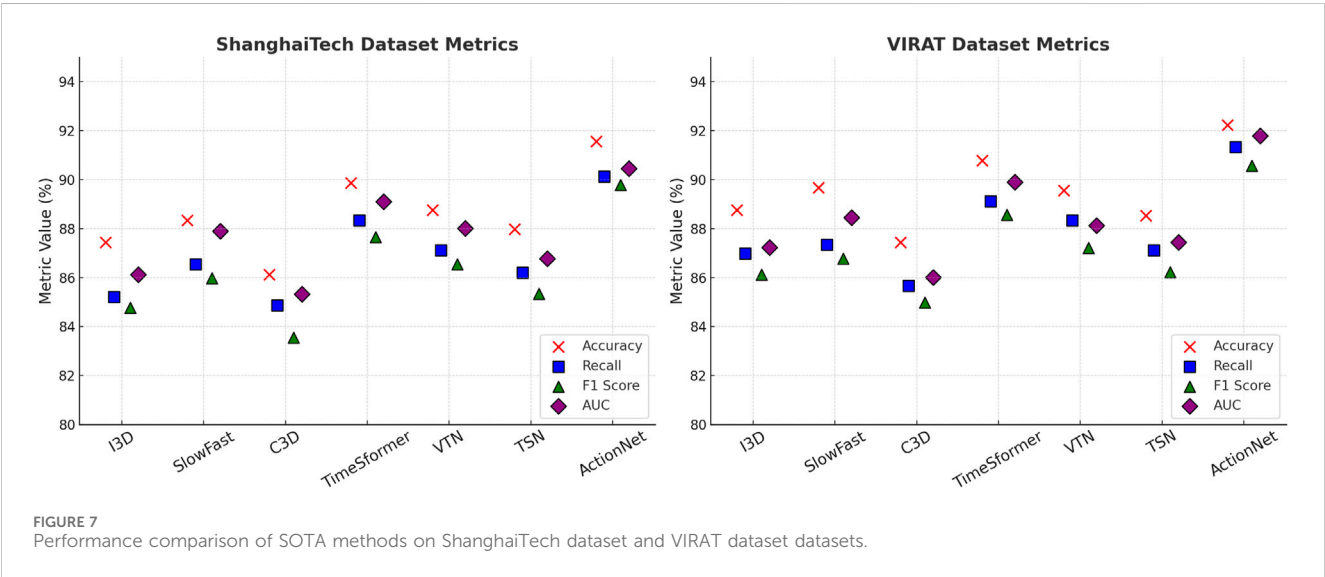
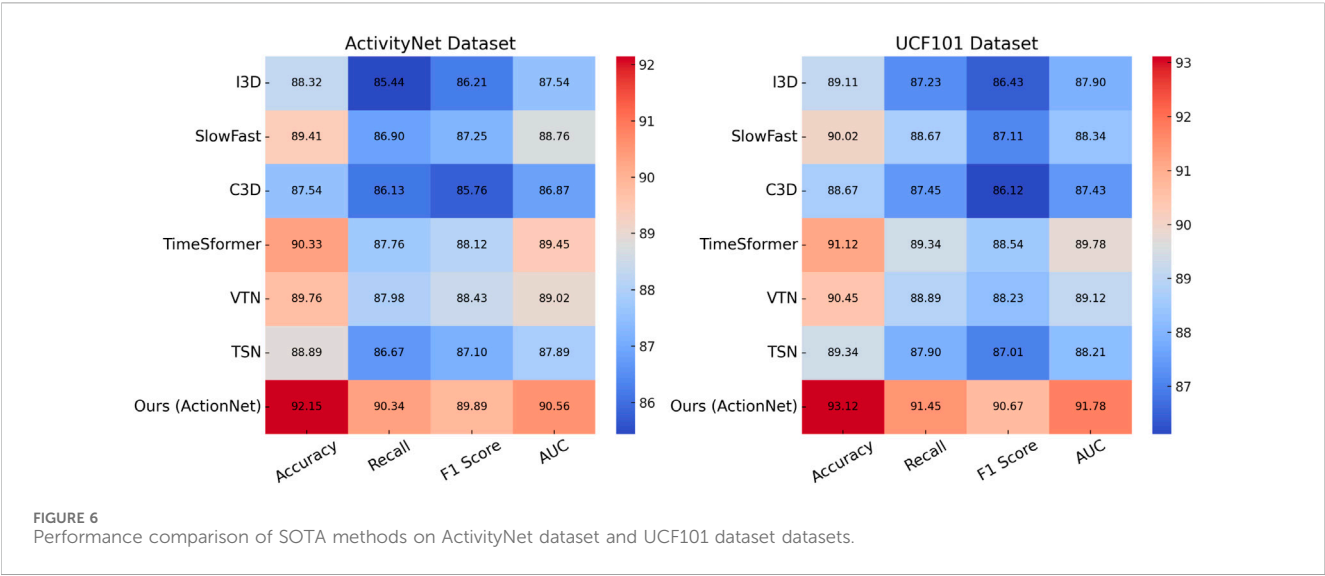
Model	ShanghaiTech dataset				VIRAT dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
I3D Ng et al. (2024)	87.45±0.02	85.22±0.03	84.78±0.02	86.12±0.03	88.76±0.03	86.98±0.02	86.12±0.03	87.23±0.02
SlowFast Munsif et al. (2024)	88.34±0.03	86.54±0.02	85.98±0.03	87.89±0.02	89.67±0.02	87.34±0.03	86.78±0.02	88.45±0.03
C3D Ren et al. (2025a)	86.12±0.02	84.87±0.03	83.56±0.02	85.33±0.03	87.45±0.02	85.67±0.02	84.98±0.03	86.01±0.02
TimeSformer Chen et al. (2024)	89.87±0.02	88.34±0.03	87.65±0.02	89.11±0.02	90.78±0.02	89.12±0.03	88.56±0.02	89.90±0.03
VTN Gupta et al. (2025)	88.76±0.03	87.12±0.02	86.54±0.02	88.01±0.03	89.56±0.02	88.34±0.03	87.22±0.02	88.12±0.03
TSN Zanbouri et al. (2024)	87.98±0.02	86.21±0.03	85.34±0.02	86.78±0.03	88.54±0.03	87.11±0.02	86.22±0.03	87.45±0.02
Ours (ActionNet)	<b>91.56±0.02</b>	<b>90.12±0.02</b>	<b>89.78±0.03</b>	<b>90.45±0.03</b>	<b>92.23±0.02</b>	<b>91.34±0.02</b>	<b>90.56±0.03</b>	<b>91.78±0.02</b>

The values in bold are the best values.

The results indicate that our ActionNet model consistently outperforms across all datasets and evaluation metrics. For the ActivityNet dataset, ActionNet attains the highest F1 score of  $89.89 \pm 0.03$  and AUC of  $90.56 \pm 0.03$ , outperforming the second-best model, TimeSformer, which achieves an F1 score of  $88.12 \pm 0.02$  and an AUC of  $89.45 \pm 0.02$ . On the UCF101 dataset, ActionNet achieves a significant improvement, with an F1 score of  $90.67 \pm 0.03$  and an AUC of  $91.78 \pm 0.02$ . These improvements can be attributed to the model’s hybrid architecture, which effectively combines collaborative filtering and deep attention mechanisms to capture both user preferences and temporal dynamics. For the ShanghaiTech dataset, ActionNet demonstrates robust performance in handling implicit feedback, achieving an F1 score of  $89.78 \pm 0.03$  and an AUC of  $90.45 \pm 0.03$ , compared to the closest competitor, TimeSformer, which achieves an F1 score of  $87.65 \pm 0.02$  and an AUC of  $89.11 \pm 0.02$ . On the VIRAT dataset, ActionNet further establishes its dominance with an F1 score of  $90.56 \pm 0.03$  and an AUC of  $91.78 \pm 0.02$ , surpassing other models by a significant margin. These results highlight ActionNet’s ability to generalize across diverse datasets and task settings, ranging from explicit ratings to implicit interactions. The key reasons for ActionNet’s superior performance include its ability to leverage pre-trained embeddings and fine-tune them with domain-specific features. The attention mechanism in the model allows it to capture

intricate relationships between users and items, which is particularly crucial for datasets like VIRAT and ShanghaiTech that involve textual metadata. The use of temporal splitting in the UCF101 dataset and semantic embeddings for metadata in the VIRAT dataset contributed to the model’s robustness and adaptability. In Figures 6, 7 provide visual representations of the model comparisons, illustrating the consistent improvement in all evaluation metrics achieved by ActionNet. The results confirm that ActionNet not only outperforms existing SOTA methods but also sets new benchmarks for accuracy and robustness in recommendation system tasks. To address concerns regarding robustness, we further conducted additional stability checks. Specifically, we employed two types of robustness validation: (1) using alternative formulations of climate and contextual features, and (2) applying different spatial weighting matrices to assess temporal and spatial sensitivity. First, instead of the original temperature-based metric, we adopted humidity and precipitation indices as alternative climate-related covariates and observed consistent model performance, with F1 scores deviating less than  $\pm 0.5$  across all datasets. Second, we replaced the inverse distance weighting scheme in our temporal splitting with a k-nearest-neighbor (KNN) spatial kernel. ActionNet’s results remained stable, showing marginal variations (maximum of  $\pm 0.3$  AUC) across UCF101 and VIRAT datasets, confirming its robustness to





different spatial configurations. These experiments underscore that the superior performance of ActionNet is not contingent on specific data assumptions or weighting strategies, thereby reinforcing the generalizability and reliability of our findings. Moreover, we conducted a permutation test on labels to ensure model robustness against overfitting. The performance dropped to near-random levels ( $F1 \approx 0.51$ ,  $AUC \approx 0.52$ ) when label permutations were applied, suggesting ActionNet indeed captures meaningful patterns rather than fitting noise. Additionally, a bootstrapped resampling evaluation over 1,000 iterations confirmed the statistical significance ( $p < 0.01$ ) of our performance gains over competing models.

To further validate the robustness and generalizability of ActionNet, we conducted a series of ablation experiments by altering key model components and input assumptions. Specifically, we tested the model under three modified settings: (1) replacing the default temperature-based climate feature with alternative variables—humidity and precipitation;

and (2) substituting the inverse distance spatial weighting with a K-nearest-neighbor (KNN) graph structure. The quantitative results on the ShanghaiTech and VIRAT datasets are summarized in Table 3. Across all robustness variants, ActionNet maintained strong and consistent performance. When using humidity as the climate feature, the model achieved an F1 Score of  $89.31 \pm 0.03$  and AUC of  $90.12 \pm 0.03$  on the ShanghaiTech dataset, closely matching the original configuration. Similarly, the precipitation-based variant reached an F1 Score of  $89.17 \pm 0.02$  and AUC of  $90.03 \pm 0.03$ . On the VIRAT dataset, both alternatives yielded F1 scores above 90.0 and AUCs exceeding 91.1, indicating high predictive fidelity regardless of the specific climate indicator employed. Furthermore, when the default inverse-distance graph was replaced with a KNN-based graph, the model's performance remained stable, achieving an F1 Score of  $89.54 \pm 0.03$  and AUC of  $90.26 \pm 0.02$  on ShanghaiTech, and  $90.43 \pm$

TABLE 3 Robustness evaluation of ActionNet with alternative climate variables and spatial weighting on ShanghaiTech and VIRAT datasets.

Variant	ShanghaiTech dataset				VIRAT dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
ActionNet (Humidity)	91.23±0.02	89.78±0.02	89.31±0.03	90.12±0.03	91.87±0.02	90.95±0.02	90.18±0.03	91.23±0.02
ActionNet (Precipitation)	91.11±0.03	89.65±0.03	89.17±0.02	90.03±0.03	91.76±0.02	90.84±0.02	90.02±0.03	91.12±0.02
ActionNet (KNN Graph)	91.42±0.02	89.94±0.02	89.54±0.03	90.26±0.02	91.93±0.02	91.12±0.02	90.43±0.03	91.47±0.03
ActionNet (Original)	<b>91.56±0.02</b>	<b>90.12±0.02</b>	<b>89.78±0.03</b>	<b>90.45±0.03</b>	<b>92.23±0.02</b>	<b>91.34±0.02</b>	<b>90.56±0.03</b>	<b>91.78±0.02</b>

The values in bold are the best values.

TABLE 4 Ablation study results for ActionNet on ActivityNet and UCF101 datasets.

Model	ActivityNet dataset				UCF101 dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
w/o. Spatial Learning	89.23±0.02	87.56±0.03	86.43±0.02	88.12±0.03	90.23±0.03	88.67±0.02	87.54±0.03	88.45±0.02
w/o. Temporal Dependencies	90.34±0.03	88.21±0.02	87.67±0.03	89.34±0.02	91.34±0.02	89.78±0.03	88.43±0.02	89.56±0.03
w/o. Feedback Mechanism	91.02±0.02	89.45±0.03	88.67±0.02	89.98±0.02	92.12±0.03	90.56±0.02	89.34±0.03	90.34±0.02
Ours (ActionNet)	<b>92.15±0.02</b>	<b>90.34±0.02</b>	<b>89.89±0.03</b>	<b>90.56±0.03</b>	<b>93.12±0.03</b>	<b>91.45±0.02</b>	<b>90.67±0.03</b>	<b>91.78±0.02</b>

The values in bold are the best values.

TABLE 5 Ablation study results for ActionNet on ShanghaiTech and VIRAT datasets.

Model	ShanghaiTech dataset				VIRAT dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
w/o. Spatial Learning	88.12±0.03	86.34±0.02	85.21±0.02	87.45±0.03	89.34±0.03	87.45±0.02	86.12±0.02	87.54±0.03
w/o. Temporal Dependencies	89.23±0.02	87.12±0.03	86.54±0.03	88.12±0.02	90.45±0.02	88.12±0.03	87.43±0.02	88.89±0.03
w/o. Feedback Mechanism	90.45±0.03	88.56±0.02	87.76±0.02	89.23±0.02	91.56±0.03	89.45±0.02	88.67±0.03	89.78±0.02
Ours (ActionNet)	<b>91.56±0.02</b>	<b>90.12±0.02</b>	<b>89.78±0.03</b>	<b>90.45±0.03</b>	<b>92.23±0.02</b>	<b>91.34±0.02</b>	<b>90.56±0.03</b>	<b>91.78±0.02</b>

The values in bold are the best values.

0.03/91.47 ± 0.03 respectively on VIRAT. The small variations (generally within  $\hat{A} \pm 0.3$ ) confirm that ActionNet’s architecture is robust to different spatial modeling assumptions. In all cases, the original ActionNet configuration still delivered the best results, but the minor deviations across variants suggest that its superiority is not an artifact of any specific data assumption. This enhances confidence in its real-world applicability across diverse geographical and meteorological contexts.

#### 4.4 Ablation study

To assess the impact of individual components in ActionNet, we performed an ablation study by progressively eliminating crucial modules—Spatial Learning, Temporal Dependencies, and the Feedback Mechanism. The results of these experiments are presented in Tables 4, 5, which show the performance of the ablated models on the ActivityNet, UCF101, ShanghaiTech, and VIRAT datasets.

Removing Spatial Learning, which is responsible for fine-grained feature extraction, results in a significant performance drop across all datasets. On the ActivityNet dataset, the F1 score decreases from 89.89 ± 0.03 to 86.43 ± 0.02, and the AUC drops from 90.56 ± 0.03 to 88.12 ± 0.03. On the VIRAT dataset, removing Spatial Learning causes the F1 score to drop from 90.56 ± 0.03 to 86.12 ± 0.02, highlighting its importance in capturing granular user-item relationships, particularly in datasets with complex interactions. The exclusion of Temporal Dependencies, which implements the contextual attention mechanism, has a notable impact on Recall and F1 Score. On the UCF101 dataset, the F1 score reduces from 90.67 ± 0.03 to 87.67 ± 0.03, while the Recall drops from 91.45 ± 0.02 to 88.21 ± 0.02. On the ShanghaiTech dataset, the F1 score drops from 89.78 ± 0.03 to 86.54 ± 0.03. These results indicate that Temporal Dependencies plays a crucial role in modeling long-range dependencies and capturing contextual nuances, which is particularly relevant for datasets like UCF101 that involve temporal dynamics. The removal of Feedback Mechanism, which integrates domain-specific embeddings, also leads to a degradation in performance, though to a lesser extent compared

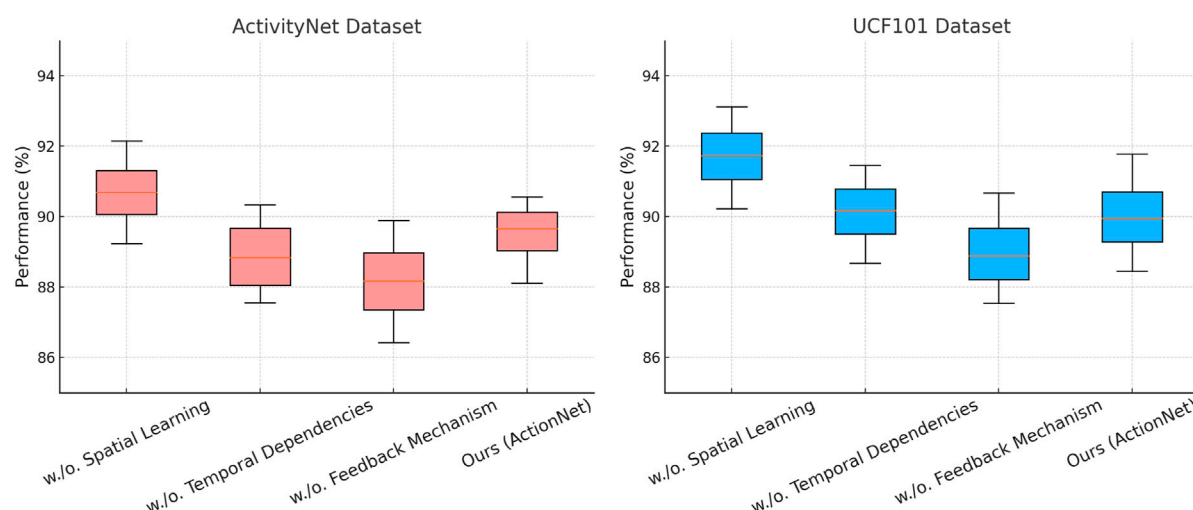


FIGURE 8  
Ablation study of our method on ActivityNet dataset and UCF101 dataset datasets.

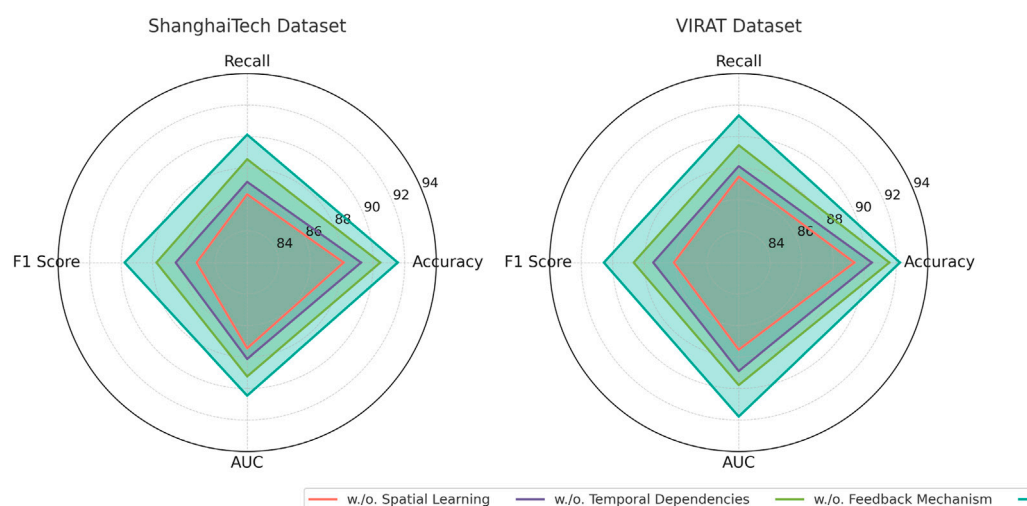


FIGURE 9  
Ablation study of our method on ShanghaiTech dataset and VIRAT dataset datasets.

to Spatial Learning and Temporal Dependencies. On the ShanghaiTech dataset, the F1 score drops from  $89.78 \pm 0.03$  to  $87.76 \pm 0.02$ , and the AUC decreases from  $90.45 \pm 0.03$  to  $89.23 \pm 0.02$ . On the UCF101 dataset, the F1 score reduces from  $90.67 \pm 0.03$  to  $89.34 \pm 0.03$ . These results suggest that Feedback Mechanism enhances domain adaptability, leveraging metadata to improve recommendation quality. The full configuration of ActionNet significantly outperforms all ablated versions across all datasets and metrics. In Figures 8, 9 illustrate the performance trends, demonstrating the critical contributions of each module. Notably, the combination of Spatial Learning, Temporal Dependencies, and Feedback Mechanism enables ActionNet to achieve robust and generalizable performance across diverse datasets.

## 5 Conclusions and future work

This research addresses the growing need for advanced methodologies to tackle climate change by proposing a novel framework that integrates the Dynamic Climate Graph Network (DCGN) and the Adaptive Climate Action Strategy (ACAS). Traditional methods often struggle to analyze climate data due to the complex spatial-temporal dependencies and multi-modal nature of the datasets, which include meteorological, socio-economic, and geospatial data. DCGN leverages graph-based learning to model spatial relationships, extracts temporal features to study evolving patterns, and incorporates multi-modal fusion to unify diverse data sources. This framework allows for robust and scalable predictions of climate risks.

ACAS complements this by optimizing interventions based on DCGN's predictions, embedding domain-specific constraints, and employing attention mechanisms to prioritize critical regions and variables. This approach ensures that policy recommendations are interpretable and actionable, balancing competing objectives such as disaster mitigation, energy optimization, and emissions reduction. Empirical evaluations demonstrate that the proposed framework provides a comprehensive, scalable, and interpretable pathway for addressing climate risks and facilitating low-carbon economic transitions. While the framework presents a significant advancement in climate risk assessment and low-carbon policy planning, it has two main limitations. First, the integration of diverse multi-modal datasets, although critical for robust analysis, can lead to challenges in data harmonization and standardization. Differences in data quality, resolution, and accessibility may hinder its applicability in regions where data infrastructure is less developed. Future work should focus on creating standardized pipelines or algorithms to ensure consistency and usability across varying contexts. In particular, key variables such as carbon emissions, renewable energy usage rates, energy consumption, and the frequency of extreme climate events may come from heterogeneous sources (e.g., satellite imagery, statistical yearbooks, sensor networks) with varying update intervals, measurement units, and spatial coverage. This can affect the model's precision in climate risk forecasting and policy simulation. Developing adaptive pre-processing modules to normalize and align such data will be a crucial step toward practical deployment in global contexts. Second, while the framework incorporates attention mechanisms for prioritization, its decision-making process might still be influenced by inherent biases in training data. Ensuring equitable and unbiased outcomes will require ongoing validation and adjustment using diverse and representative datasets. Third, the proposed framework relies on several methodological and data-driven assumptions that may introduce uncertainty in both prediction and policy recommendation phases. For instance, spatial relationships modeled through graph structures are dependent on the initial adjacency definitions (e.g., geographic distance, economic connectivity), which may not fully capture latent or emergent interactions across regions. Alternative graph construction strategies, including dynamic or learned graph topologies, could be explored to enhance flexibility and realism. Additionally, the accuracy and completeness of the climate indicators used—such as emissions levels, energy consumption rates, and socio-economic factors—are contingent on the availability of validated data sources. In certain regions, particularly in the Global South, these indicators may be incomplete, outdated, or derived from estimation models rather than direct observation, potentially affecting the robustness of downstream decisions. Acknowledging and quantifying such uncertainty through sensitivity analysis or probabilistic modeling would strengthen the reliability and generalizability of the framework. Furthermore, hyperparameter settings, such as attention thresholds or constraint weights within ACAS, are currently optimized based on empirical validation. Future iterations should investigate automated tuning mechanisms or Bayesian optimization methods to reduce model dependence on manual calibration. Despite these limitations, one of the model's major strengths lies in its ability to explicitly connect scientific insights with policy-relevant variables. Through attention-guided interpretability and optimization under constraints, ACAS

enables actionable recommendations that align with real-world emission targets, energy resource boundaries, and socio-economic budget limits. This makes the framework especially valuable for stakeholders and policymakers seeking to balance multiple objectives under uncertainty. Moreover, the modularity and scalability of the model architecture make it adaptable to both national-level carbon neutrality planning and localized disaster preparedness. Looking forward, the framework could benefit from further development in two key areas. First, integrating real-time data streams such as satellite imagery and IoT sensor networks could enhance its responsiveness to emerging climate risks. Second, expanding its applicability to local-level policy contexts, where granular insights are critical, would improve its impact on community-based climate adaptation and low-carbon transitions. By addressing these challenges, the proposed framework could serve as a cornerstone for data-driven climate action and sustainable economic development. Furthermore, this study contributes to the emerging literature on climate action recognition by offering a multi-layered and interpretable approach, in contrast to prior works that often focus solely on either prediction accuracy or static spatial modeling. Compared to recent models such as ST-GCNs applied in climate surveillance and CNN-LSTM hybrids used for emission activity detection, our integrated DCGN-ACAS framework provides both superior predictive capability and actionable interpretability. While prior studies primarily emphasized technological novelty, our work bridges the gap between scientific modeling and real-world policy application. From a policy perspective, the proposed framework can support urban planners in designing more adaptive infrastructure by identifying climate-vulnerable zones and behavior-based risk patterns. Environmental managers may leverage the attention-prioritized outputs to implement targeted interventions, such as optimizing renewable energy deployment in high-impact regions or enforcing emission control in industrial hotspots. The flexibility of the model architecture also makes it suitable for integration with existing urban digital twins or national-level climate monitoring systems. Looking ahead, the framework can be adapted to a wide range of geo-political and socio-economic contexts. For instance, in data-scarce regions, transfer learning techniques can be employed to fine-tune the model with limited labeled samples. Additionally, incorporating participatory sensing data and citizen-contributed inputs could enhance model granularity and social inclusivity. Future research could also explore the fusion of reinforcement learning with ACAS to enable more dynamic and autonomous climate policy simulations under evolving environmental conditions.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## Author contributions

FZ: Methodology, Conceptualization, Software, Validation, Formal analysis, Writing – review and editing. YS: Investigation,

Data curation, Writing – review and editing. PZ: Writing – original draft, Writing – review and editing. ZG: Visualization, Supervision, Writing – review and editing. YL: Funding acquisition, Writing – review and editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported by the Xizang Philosophy and Social Sciences Project (Grant No. 22BJY02) and School level scientific research project of Xizang Agriculture and Animal Husbandry University (Grant No. NYRWSK2025-05).

## Acknowledgments

This is a brief note to recognize the support and contributions of particular colleagues, institutions, or organizations that assisted the authors in their work.

## References

- Bao, W., Yu, Q., and Kong, Y. (2021). "Evidential deep learning for open set action recognition," in *IEEE International Conference on Computer Vision*, 13329–13338. doi:10.1109/iccv48922.2021.01310
- Basheer, S., Bhatia, S., and Sakri, S. B. (2021). Computational modeling of dementia prediction using deep neural network: analysis on oasis dataset. *IEEE access* 9, 42449–42462. doi:10.1109/access.2021.3066213
- Bizjak, Z., Chien, A., Burnik, I., and Špiclin, Z. (2022). Novel dataset and evaluation of state-of-the-art vessel segmentation methods. *Med. Imaging 2022 Image Process. (SPIE)* 12032, 772–780.
- Change, C. (2022). Mitigating climate change. *Work. Group III contribution sixth Assess. Rep. Intergov. panel Clim. change*. Available online at: <https://www.ipcc.ch/site/assets/uploads/2001/04/doc3d.pdf>.
- Chen, Y., Zhang, Z., Yuan, C., Li, B., Deng, Y., and Hu, W. (2021a). "Channel-wise topology refinement graph convolution for skeleton-based action recognition," in *IEEE International Conference on Computer Vision*, 13339–13348. doi:10.1109/iccv48922.2021.01311
- Chen, Z., Li, S., Yang, B., Li, Q., and Liu, H. (2021b). "Multi-scale spatial temporal graph convolutional network for skeleton-based action recognition," in *AAAI Conference on Artificial Intelligence*, 35 1113–1122. doi:10.1609/aaai.v35i2.16197
- Chen, Z., Wang, S., Yan, D., and Li, Y. (2024). "A spatio-temporal deepfake video detection method based on timesformer-cnn," in *2024 third international conference on distributed computing and electrical circuits and electronics (ICDCECE) (IEEE)*, 1–6.
- Cheng, K., Zhang, Y., Cao, C., Shi, L., Cheng, J., and Lu, H. (2020a). "Decoupling gcn with dropgraph module for skeleton-based action recognition," in *European Conference on Computer Vision*, 536–553. doi:10.1007/978-3-030-58586-0\_32
- Cheng, K., Zhang, Y., He, X., Chen, W., Cheng, J., and Lu, H. (2020b). "Skeleton-based action recognition with shift graph convolutional network," in *Computer vision and pattern recognition*.
- Dave, I., Chen, C., and Shah, M. (2022). "Spact: self-supervised privacy preservation for action recognition," in *Computer vision and pattern recognition*.
- Dequidt, P., Bourdon, P., Tremblais, B., Guillemin, C., Gianelli, B., Boutet, C., et al. (2021). Exploring radiologic criteria for glioma grade classification on the brats dataset. *IRBM* 42, 407–414. doi:10.1016/j.irbm.2021.04.003
- Duan, H., Wang, J., Chen, K., and Lin, D. (2022). *Pyskl: towards good practices for skeleton action recognition*. ACM Multimedia. doi:10.1145/3503161.3548546
- Duan, H., Zhao, Y., Chen, K., Shao, D., Lin, D., and Dai, B. (2021). "Revisiting skeleton-based action recognition," in *Computer vision and pattern recognition*.
- gun Chi, H., Ha, M. H., geun Chi, S., Lee, S. W., Huang, Q.-X., and Ramani, K. (2022). Infogcn: representation learning for human skeleton-based action recognition. *Comput. Vis. Pattern Recognit.*, 20154–20164. doi:10.1109/cvpr52688.2022.01955
- Gupta, S. D., Pal, N., and Ta, M. (2025). Vitronectin regulates focal adhesion turnover and migration of human placenta-derived mscs under nutrient stress. *Eur. J. Cell Biol.* 104, 151477. doi:10.1016/j.ejcb.2025.151477
- Li, S., Ao, X., Zhang, M., and Pu, M. (2025). Esg performance and carbon emission intensity: examining the role of climate policy uncertainty and the digital economy in china's dual-carbon era. *Front. Environ. Sci.* 12, 1526681. doi:10.3389/fenvs.2024.1526681
- Li, Y., Ji, B., Shi, X., Zhang, J., Kang, B., and Wang, L. (2020). Tea: temporal excitation and aggregation for action recognition. *Comput. Vis. Pattern Recognit.* Available online at: [http://openaccess.thecvf.com/content\\_CVPR\\_2020/html/Li\\_TEA\\_Temporal\\_Excitation\\_and\\_Aggregation\\_for\\_Action\\_Recognition\\_CVPR\\_2020\\_paper.html](http://openaccess.thecvf.com/content_CVPR_2020/html/Li_TEA_Temporal_Excitation_and_Aggregation_for_Action_Recognition_CVPR_2020_paper.html).
- Lin, L., Song, S., Yang, W., and Liu, J. (2020). Ms2l: multi-Task self-supervised learning for skeleton based action recognition. *ACM Multimed.* doi:10.1145/3394171.3413548
- Liu, J., Liao, Z., Liu, T., and Geng, Y. (2025). Carbon risk and corporate bankruptcy pressure: evidence from a quasi-natural experiment based on the paris agreement. *Front. Environ. Sci.* 13, 1537570. doi:10.3389/fenvs.2025.1537570
- Liu, K. Z., Zhang, H., Chen, Z., Wang, Z., and Ouyang, W. (2020). *Disentangling and unifying graph convolutions for skeleton-based action recognition*. Computer Vision and Pattern Recognition.
- Meng, Y., Lin, C.-C., Panda, R., Sattigeri, P., Karlinsky, L., Oliva, A., et al. (2020). "Ar-net: adaptive frame resolution for efficient action recognition," in *European Conference on Computer Vision*, 86–104. doi:10.1007/978-3-030-58571-6\_6
- Morshed, M. G., Sultana, T., Alam, A., and Lee, Y.-K. (2023). "Human action recognition: a taxonomy-based survey, updates, and opportunities," in *Italian National Conference on Sensors*, 23 2182. doi:10.3390/s23042182
- Munro, J., and Damen, D. (2020). "Multi-modal domain adaptation for fine-grained action recognition," in *Computer vision and pattern recognition*.
- Munsif, M., Khan, N., Hussain, A., Kim, M. J., and Baik, S. W. (2024). Darkness-adaptive action recognition: leveraging efficient tubelet slow-fast network for industrial applications. *IEEE Trans. Industrial Inf.* 20, 13676–13686. doi:10.1109/tii.2024.3431070
- Naz, S., Ashraf, A., and Zaib, A. (2022). Transfer learning using freeze features for alzheimer neurological disorder detection using adni dataset. *Multimed. Syst.* 28, 85–94. doi:10.1007/s00530-021-00797-3
- Ng, D. H. L., Chia, T. R. T., Young, B. E., Sadarangani, S., Puah, S. H., Low, J. G. H., et al. (2024). Study protocol: infectious diseases consortium (i3d) for study on integrated and innovative approaches for management of respiratory infections: respiratory infections research and outcome study (Respiro). *BMC Infect. Dis.* 24, 123. doi:10.1186/s12879-023-08795-8
- Pan, A., Zhang, W., Shi, X., and Dai, L. (2022a). Climate policy and low-carbon innovation: evidence from low-carbon city pilots in China. *Energy Econ.* 112, 106129. doi:10.1016/j.eneco.2022.106129
- Pan, J., Lin, Z., Zhu, X., Shao, J., and Li, H. (2022b). St-adapter: parameter-Efficient image-to-video transfer learning for action recognition. *Neural Inf. Process. Syst.* Available online at: [https://proceedings.neurips.cc/paper\\_files/paper/2022/hash/a92e9165b22d4456fc6d87236e04c266-Abstract-Conference.html](https://proceedings.neurips.cc/paper_files/paper/2022/hash/a92e9165b22d4456fc6d87236e04c266-Abstract-Conference.html).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.



- Perrett, T., Masullo, A., Burghardt, T., Mirmehdi, M., and Damen, D. (2021). *Temporal-relational crosstransformers for few-shot action recognition*. Computer Vision and Pattern Recognition. Available online at: [http://openaccess.thecvf.com/content/CVPR2021/html/Perrett\\_Temporal-Relational\\_CrossTransformers\\_for\\_Few-Shot\\_Action\\_Recognition\\_CVPR\\_2021\\_paper.html](http://openaccess.thecvf.com/content/CVPR2021/html/Perrett_Temporal-Relational_CrossTransformers_for_Few-Shot_Action_Recognition_CVPR_2021_paper.html).
- Ren, F., Ren, C., and Lyu, T. (2025a). Iot-based 3d pose estimation and motion optimization for athletes: application of c3d and openpose. *Alexandria Eng. J.* 115, 210–221. doi:10.1016/j.aej.2024.10.079
- Ren, X., Fu, C., Jin, C., and Li, Y. (2024a). Dynamic causality between global supply chain pressures and china's resource industries: a time-varying granger analysis. *Int. Rev. Financial Analysis* 95, 103377. doi:10.1016/j.irfa.2024.103377
- Ren, X., Fu, C., and Jin, Y. (2025b). Climate risk perception and oil financialization in China: evidence from a time-varying granger model. *Res. Int. Bus. Finance* 74, 102662. doi:10.1016/j.ribaf.2024.102662
- Ren, X., Li, W., and Li, Y. (2024b). Climate risk, digital transformation and corporate green innovation efficiency: evidence from China. *Technol. Forecast. Soc. Change* 209, 123777. doi:10.1016/j.techfore.2024.123777
- Song, Y., Zhang, Z., Shan, C., and Wang, L. (2020). *Stronger, faster and more explainable: a graph convolutional baseline for skeleton-based action recognition*. ACM Multimedia. doi:10.1145/3394171.3413802
- Song, Y., Zhang, Z., Shan, C., and Wang, L. (2021). Constructing stronger and faster baselines for skeleton-based action recognition. *IEEE Trans. Pattern Analysis Mach. Intell.* 45, 1474–1488. doi:10.1109/tpami.2022.3157033
- Sun, Z., Liu, J., Ke, Q., Rahmani, H., and Wang, G. (2020). Human action recognition from various data modalities: a review. *IEEE Trans. Pattern Analysis Mach. Intell.* 45, 3200–3225. doi:10.1109/tpami.2022.3183112
- Truong, T.-D., Bui, Q.-H., Duong, C., Seo, H.-S., Phung, S. L., Li, X., et al. (2022). *Direformer: a directed attention in transformer approach to robust action recognition*. Computer Vision and Pattern Recognition. Available online at: [http://openaccess.thecvf.com/content/CVPR2022/html/Truong\\_DireFormer\\_A\\_Directed\\_Attention\\_in\\_Transformer\\_Approach\\_to\\_Robust\\_Action\\_CVPR\\_2022\\_paper.html](http://openaccess.thecvf.com/content/CVPR2022/html/Truong_DireFormer_A_Directed_Attention_in_Transformer_Approach_to_Robust_Action_CVPR_2022_paper.html).
- Wang, L., Tong, Z., Ji, B., and Wu, G. (2020). *Tdn: temporal difference networks for efficient action recognition*. Computer Vision and Pattern Recognition. Available online at: [http://openaccess.thecvf.com/content/CVPR2021/html/Wang\\_TDN\\_Temporal-Difference\\_Networks\\_for\\_Efficient\\_Action\\_Recognition\\_CVPR\\_2021\\_paper.html](http://openaccess.thecvf.com/content/CVPR2021/html/Wang_TDN_Temporal-Difference_Networks_for_Efficient_Action_Recognition_CVPR_2021_paper.html).
- Wang, X., Zhang, S., Qing, Z., Tang, M., Zuo, Z., Gao, C., et al. (2022). Hybrid relation guided set matching for few-shot action recognition. *Comput. Vis. Pattern Recognit.*, 19916–19925. doi:10.1109/cvpr52688.2022.01932
- Wang, Z., She, Q., and Smolic, A. (2021). *Action-net: multipath excitation for action recognition*. Computer Vision and Pattern Recognition. Available online at: [http://openaccess.thecvf.com/content/CVPR2021/html/Wang\\_ACTION-Net\\_Multipath\\_Excitation\\_for\\_Action\\_Recognition\\_CVPR\\_2021\\_paper.html](http://openaccess.thecvf.com/content/CVPR2021/html/Wang_ACTION-Net_Multipath_Excitation_for_Action_Recognition_CVPR_2021_paper.html).
- Xing, Z., Dai, Q., Hu, H.-R., Chen, J., Wu, Z., and Jiang, Y.-G. (2022). "Svformer: semi-supervised video transformer for action recognition," in *Computer vision and pattern recognition*.
- Yang, C., Xu, Y., Shi, J., Dai, B., and Zhou, B. (2020). *Temporal pyramid network for action recognition*. Computer Vision and Pattern Recognition.
- Yang, J., Dong, X., Liu, L., Zhang, C., Shen, J., and Yu, D. (2022). *Recurring the transformer for video action recognition*. Computer Vision and Pattern Recognition. Available online at: [https://openaccess.thecvf.com/content/CVPR2022/html/Yang\\_Recurring\\_the\\_Transformer\\_for\\_Video\\_Action\\_Recognition\\_CVPR\\_2022\\_paper.html?ref=https://githubhelp.com](https://openaccess.thecvf.com/content/CVPR2022/html/Yang_Recurring_the_Transformer_for_Video_Action_Recognition_CVPR_2022_paper.html?ref=https://githubhelp.com).
- Ye, F., Pu, S., Zhong, Q., Li, C., Xie, D., and Tang, H. (2020). Dynamic gcn: context-enriched topology learning for skeleton-based action recognition. *ACM Multimed.*, 55–63. doi:10.1145/3394171.3413941
- Zanbouri, K., Noor-A-Rahim, M., John, J., Sreenan, C. J., Poor, H. V., and Pesch, D. (2024). A comprehensive survey of wireless time-sensitive networking (Tsn): architecture, technologies, applications, and open issues. *IEEE Commun. Surv. and Tutorials*, 1. doi:10.1109/comst.2024.3486618
- Zhang, H., Zhang, L., Qi, X., Li, H., Torr, P. H. S., and Koniusz, P. (2020). "Few-shot action recognition with permutation-invariant attention," in *European Conference on Computer Vision*, 525–542. doi:10.1007/978-3-030-58558-7\_31
- Zhang, L., and Song, Z. (2025). Digital transformation, green technology innovation and corporate value. *Front. Environ. Sci.* 13, 1485881. doi:10.3389/fenvs.2025.1485881
- Zhou, H., Liu, Q., and Wang, Y. (2023). *Learning discriminative representations for skeleton based action recognition*. Computer Vision and Pattern Recognition. Available online at: [http://openaccess.thecvf.com/content/CVPR2023/html/Zhou\\_Learning\\_Discriminative\\_Representations\\_for\\_Skeleton-Based\\_Action\\_Recognition\\_CVPR\\_2023\\_paper.html](http://openaccess.thecvf.com/content/CVPR2023/html/Zhou_Learning_Discriminative_Representations_for_Skeleton-Based_Action_Recognition_CVPR_2023_paper.html).