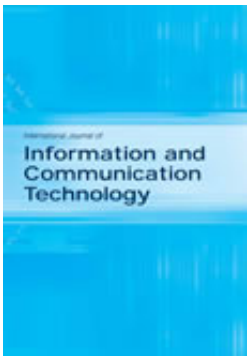


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Spatiotemporal evolution analysis of extreme weather based on adaptive graph convolutional networks

Yadi Zhang, Lin Zhou, Wencheng Sun

DOI: 10.1504/IJICT.2025.10074941

Article History:

Received:	01 September 2025
Last revised:	02 October 2025
Accepted:	02 October 2025
Published online:	17 December 2025

Spatiotemporal evolution analysis of extreme weather based on adaptive graph convolutional networks

Yadi Zhang*, Lin Zhou and Wencheng Sun

Southwest Branch,
State Grid Corporation of China,
Chengdu, 610040, China
Email: yadizwh@126.com
Email: zhoulin@sw.sgcc.com.cn
Email: yanxc126@126.com
*Corresponding author

Abstract: Extreme weather is intensifying and demands forecasts that capture how hazards emerge and travel. To address non-stationary dependencies and rare tails, this paper introduces an adaptive spatiotemporal graph framework. In the scheme, first a fused relational graph blends physical priors with a learned adjacency to follow changing pathways; then a multi-scale temporal encoder links weak precursors to rapid surges; finally calibrated decision layers set cost-aware thresholds and deliver coherent maps with an audit trail. Experiments show gains at six-hour lead: area under the precision-recall curve rises by 8.5% over graph WaveNet and 22% over LSTM, fractions skill at ten kilometres improves from 0.44 to 0.52, the Brier score drops from 0.134 to 0.117, and false alarms fall by about ten percent, while seasonal scores remain between 0.41 and 0.53.

Keywords: adaptive graph convolution; extreme weather; spatiotemporal dependence; teleconnections; early warning.

Reference to this paper should be made as follows: Zhang, Y., Zhou, L. and Sun, W. (2025) 'Spatiotemporal evolution analysis of extreme weather based on adaptive graph convolutional networks', *Int. J. Information and Communication Technology*, Vol. 26, No. 46, pp.95–113.

Biographical notes: Yadi Zhang received his PhD from Wuhan University, currently serves as a senior engineer at the Southwest Branch of State Grid Corporation of China. His research interests include high-voltage insulation technology, new technologies for transmission operation, engineering disaster prevention, and power meteorology.

Lin Zhou received a Bachelor's degree from Sichuan University of Science and Engineering, currently serves as a senior engineer at the Southwest Branch of State Grid Corporation of China. His research interests include high-voltage insulation technology, new technologies for transmission operation, and engineering disaster prevention.

Wencheng Sun received a Master's degree from Sichuan University, currently works as a senior engineer at the Southwest Branch of State Grid Corporation of China. His research interests include high-voltage insulation technology, new technologies for transmission operation, and engineering disaster prevention.

1 Introduction

By the early 21st century, extreme weather shifted from statistical curiosities to constraints that shape planning in disaster response, farming, energy systems, and public health. Heat that stalls for days over one basin, convective bursts that drown streets before pumps start, and compound episodes where wind, surge, and rain arrive together now demand explanations that follow an event's birth, drift, and decay through space and time (Apicella et al., 2023). Building such explanations is hard. Extremes live in the thinnest parts of the record, yet the forces that prime and steer them range from boundary-layer feedbacks and mesoscale organisation to basin-scale moisture transport and planetary-scale teleconnections. Observations come as an untidy weave of station logs, satellite swaths, and reanalysis grids, each with its own blind spots and noise (Chen et al., 2024). Numerical prediction remains indispensable but is strained by resolution limits, uncertain parameterisations, and regime-dependent bias just where nonlinearity bites. Classical statistics summarise tails and return levels and are valuable for risk framing, but they rarely describe how an event's footprint deforms as it crosses mountains, rivers, or dense urban fabric (Fu et al., 2022). Data-driven models promised relief by mining large archives, and they deliver useful texture and persistence (Gao et al., 2022). Yet most treat space as a fixed Euclidean lattice and assume that dependence patterns do not move. The atmosphere refuses that convenience. Distant links strengthen or fade with season and background flow; local ties tighten along fronts and slacken as structures decay. Systems that cannot change their notion of 'who is a neighbour' in step with regime shifts miss early signals, displace peaks, and produce confident ghosts (Hu et al., 2022). A more faithful approach views geophysical fields as relational systems whose connectivity is partly prescribed by geography and physics and partly discovered from data – and crucially, allowed to evolve in time. Such a view makes room for enduring pathways like orographic channels and for transient corridors such as event-locked moisture streams, bringing predictive models closer to the way forecasters reason about real storms (Hu et al., 2022).

Graph-based learning offers a principled path toward that representation. By assigning nodes to stations or grid cells and edges to relationships such as proximity, shared watershed, correlated variability, or prevailing flow alignment, a graph encodes the scaffolding through which disturbances travel (Jiang and Deng, 2024). Spatiotemporal graph networks can then perform message passing that respects this scaffolding while aggregating information across time, capturing both local contagion and far-field influence. Progress to date has demonstrated that static graphs derived from distance or climatological correlation improve over purely grid-convolutional baselines for routine phenomena. Extremes, however, demand adaptivity (Korban et al., 2023). During a heatwave, the effective connectivity between regions may strengthen along synoptic ridges and weaken across maritime boundaries; during a flood episode, orographic lifting can create directional pathways that are absent in quiescent periods. Teleconnections linked to background indices shift nodes that are 'neighbours' in a relational sense, even if they remain far apart geographically (Li et al., 2024). A model that holds its adjacency fixed implicitly assumes that the routes of influence are time-invariant, which mischaracterises the true, regime-dependent flow of information. Models that rely on a single, fixed notion of neighbourhood struggle when atmospheric linkages rearrange under shifting flow. A stronger choice is to let connectivity be learned from data and to anchor it with physical priors so the resulting graph remains

interpretable and resilient to noise. Prior knowledge firmly anchors structural relationships to geography, terrain, and dominant transport mechanisms, while learning components reveal transient pathways such as moisture plumes that occur only during specific seasonal phases (Liu et al., 2021). The absence of anchor points leads to shortcut learning and excessive sensitivity to noise, whereas relying solely on prior knowledge overlooks non-stationary patterns and rare couplings that manifest under pressure conditions (Liu et al., 2024). Therefore, careful orchestration is required: static scaffolding ensures fundamental geographic and physical consistency; adaptive layers capture context-dependent relationships; and explicit constraints-sparse, normalised, and stability penalties-ensure learned maps remain applicable beyond storm training. Temporal complexity further intensifies: upstream precursors may emerge days in advance, while local thresholds can abruptly trigger within hours, requiring encoders to balance long-term coverage with abrupt changes (Ma et al., 2023). Holes in causal convolution provide long-torque receptive fields without compromising temporal causality; streamlined attention mechanisms highlight critical lagging factors; gated updates preserve phase information and prevent signal attenuation. Given the archival data's predominantly calm periods compared to extreme events, training must strengthen recall at decision-relevant thresholds while avoiding false alarm spikes. Credibility demands uncompromising rigor: scores undergo calibration to determine operational parameters, with interpreted results mapping alert nodes and edges for forecasters to cross-reference model logic with weather pattern analysis (Pan et al., 2022).

Based on this research, we constructed an adaptive graph framework for spatiotemporal evolution of extreme events. The graph is viewed as a 'time-aware' composite system: it is built upon robust priors derived from distance, terrain, and typical circulation patterns, supplemented by data-conditioned layers reconstructed dynamically. Candidate edge connections are generated through node embedding refinement from recent history, with their strengths normalised and trimmed to retain only credible pathways. Message passing through this fusion operator enables both persistent and emerging relationships to coexist. Temporal dynamics are processed via multi-scale stacking: extended receptive fields capture long-term dependencies, while lightweight attention mechanisms refine distant influences when necessary (Peng et al., 2023). Slow-changing drivers-climate indices, soil and underlying surface descriptors-are injected as conditional variables to guide the encoder in producing context-consistent interpretations. The objective function balances category detection with intensity characterisation, incorporating smoothness and stability terms to prevent structural 'fluctuations' across batches (Wei et al., 2023). During validation and testing phases, probabilities are calibrated and converted into thresholds based on error-tolerance parameters. Evaluation assesses not only discriminative power and reliability but also spatial skill and trajectory fidelity: verifying whether predicted extremes fall within correct pathways, and whether growth/decay patterns align with observations. The framework simultaneously outputs auditable outputs: timestamped adjacency maps display activated cross-correlations, contribution profiles map alarm-triggering linkages upstream, and migration path visualisations visualise the movement trajectories of hazardous zones. Hydrologists can see which tributaries signalled rising flood risk, city teams can check whether predicted heat islands align with vulnerable districts, and forecasters can judge whether the system recognised a blocking pattern or merely extrapolated local persistence (Xia et al., 2024).

2 Relevant technologies

2.1 Scientific and operational background

Extreme weather is better understood as a moving sequence than as a simple spike in temperature or rainfall. Heat can settle over a basin under subsidence and cloudless skies, then shift as ridge patterns strengthen or break. Flooding rain tends to organise along moisture pathways carved by low-level jets, upslope lifting, and mesoscale convective lines. These patterns seldom honour fixed neighbourhoods on a map (Xiong et al., 2024). Cells merge, edges sharpen, and far-reaching linkages wax or fade with the season. Any useful model must account for three facts at once: behaviour in the extremes rather than the average, dependence that drifts across space and time, and external drivers that redirect communication between regions (Yang et al., 2022).

The evidence base carries the same complexity. Station networks provide long and trusted records at specific points, yet coverage is uneven. Satellites offer broad views and rapid revisits, but retrievals vary with surface type and cloud structure. Reanalysis ties fields together dynamically, though it inherits biases from model physics and assimilation choices. Bringing these sources into a single frame calls for disciplined preparation: align time steps to a common cadence, register grids to remove reprojection and geolocation drift, and fill gaps with methods that keep peaks intact instead of smoothing them away. Event labels must be explicit and defensible, such as percentile rules for heat or exceedance methods for heavy rain that respect local climatology (Yang et al., 2023).

Practical constraints tighten the problem. By design, extremes are scarce, so archives are dominated by quiet periods. A naïve classifier can post a high overall score simply by saying ‘no event’ most of the time, while missing the very cases that matter. Effective learning therefore emphasises recall at decision-relevant thresholds, tempers false alarms through cost-aware choices, and reports calibrated probabilities so operators can judge confidence rather than rely on sharp but unreliable scores (Zang et al., 2024). False alarms also carry social and economic costs. Decision makers need calibrated scores that align with their tolerance for misses and over-warnings. Spatial verification is equally important. A forecast that places a flood risk 20 grid cells away is not operationally useful, no matter how accurate its aggregate statistics. These conditions motivate models that integrate imbalance-aware learning, probabilistic calibration, and spatially explicit skill assessment.

2.2 Statistical foundations for extremes and dependence

Classical extreme value theory provides a language for marginal tail behaviour. Two families are widely used in geosciences. The block maxima approach models the maximum within fixed windows such as months or years. The peak-over-threshold approach models exceedances above a high threshold to make better use of rare information. For block maxima, the generalised extreme value distribution is a common choice:

$$G(z) = \exp\left\{-\left[1+\xi\frac{z-\mu}{\sigma}\right]^{-1/\xi}\right\}, \quad 1+\xi\frac{z-\mu}{\sigma} > 0 \tag{1}$$

where z is the block maximum, μ is the location parameter that shifts the distribution, $\sigma > 0$ is the scale parameter that stretches or compresses the tail, and ξ is the shape parameter that controls tail heaviness and distinguishes between bounded, light, and heavy tails.

While such parametric forms summarise severity and return levels, extremes rarely act independently in space or time. Tail dependence may persist over long distances because teleconnections alter the effective geometry of the atmosphere. A stationary, Euclidean assumption that neighbor cells are those separated by the smallest physical distance often underestimates these couplings. Several strategies attempt to bridge the gap. Copula-based models encode multivariate dependence beyond linear correlation and can target tail dependence directly. Max-stable processes extend extreme value theory to continuous fields but entail considerable computational cost for high-resolution grids. Hierarchical Bayesian schemes combine process priors with data-driven updates and allow covariates such as topography or climate indices to shape spatial structure. Despite their strengths, these approaches can struggle to represent rapidly changing connectivity during an unfolding event (Zhang et al., 2022).

Graph representations offer an alternative that aligns with the relational nature of atmospheric dynamics. Nodes denote stations or grid cells. Edges denote a relation such as distance-based affinity, similarity in historical anomalies, common watershed membership, or alignment with typical flow directions. The graph becomes a scaffolding that constrains how information passes from one node to another. This perspective supports a flexible synthesis of prior knowledge and learned interactions. A watershed graph captures hydrologic pathways for flood propagation. A k-nearest-neighbour graph imposes minimal structure when prior knowledge is sparse. A correlation graph learned from historical seasons encodes typical teleconnections. Crucially, graphs can be made adaptive so that edge strengths depend on the current state rather than fixed climatology (Zhang et al., 2023).

Dependence in time requires equal care. Extremes can be preceded by subtle precursors at long lead times while the final escalation happens quickly. Standard autoregressive models capture local persistence but miss multiple scales of delay. Temporal convolution with dilation builds large receptive fields without collapsing sequence order. Attention mechanisms highlight scattered yet influential lags. Recurrent gates preserve phase information and suppress vanishing gradients. No single mechanism solves the entire problem; instead, modern systems often blend these encoders to respect both slow-building regimes and abrupt transitions.

2.3 Graph learning and convolutional operators

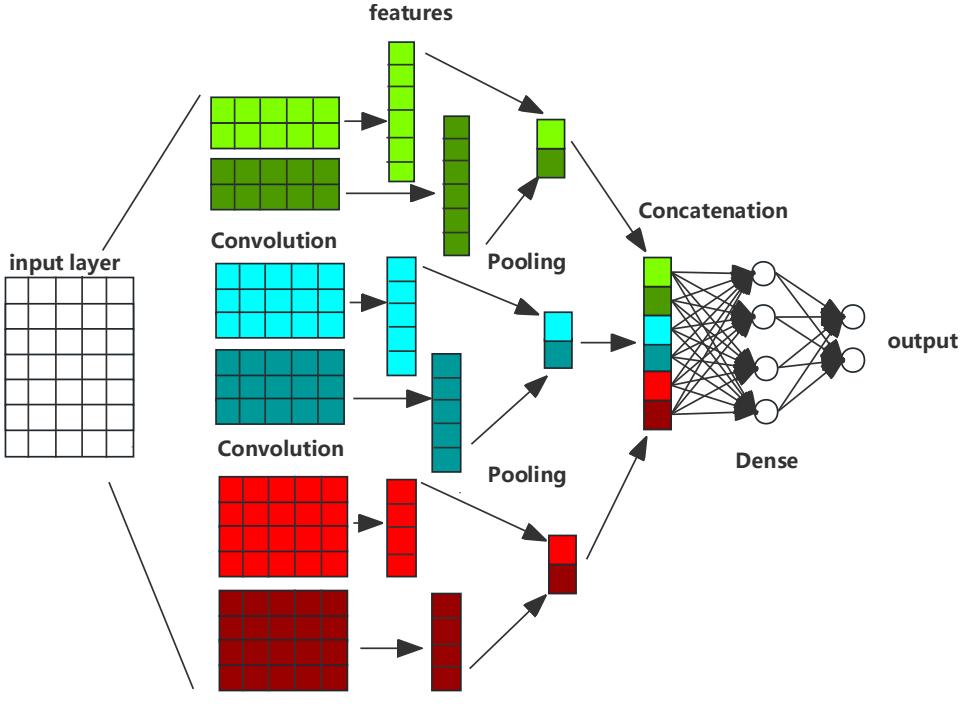
Graph-based deep learning provides a general recipe for reasoning on irregular domains. The central operation is message passing: each node aggregates information from its neighbours, transforms it through learned weights, and updates its state. A widely used formulation of graph convolution employs a normalised adjacency to stabilise learning:

$$H^{(l+1)} = \sigma\left(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \tag{2}$$

where $H^{(l)}$ is the node-feature matrix at layer l, $\hat{A} = A + I$ is the adjacency with self-loops, A is the base adjacency, I is the identity matrix that preserves node-specific

information, $\hat{D}$ is the diagonal degree matrix of $\hat{A}$, $W^{(l)}$ is a learnable weight matrix, and σ is a nonlinear activation that introduces capacity beyond linear propagation. To benchmark graph-based methods, a conventional Euclidean CNN pipeline is considered, as illustrated in Figure 1, which extracts spatial features from gridded reanalysis before dense prediction.

Figure 1 Canonical CNN pipeline for grid-based fields (see online version for colours)



The power of this operator rests on the choice of A. A purely distance-based A respects geography but ignores regime-dependent couplings. A purely data-driven A can discover hidden pathways but risks amplifying spurious correlations. Several design patterns address this tension (Zhao et al., 2022).

1 Static priors with adaptive refinement. A static component encodes minimum structure such as geodesic proximity, basin connectivity, or prevailing wind corridors. An adaptive component is learned in an embedding space where nodes acquire representations that summarise recent dynamics. The adaptive adjacency arises from projected similarities of these embeddings and is normalised and sparsified to maintain numerical stability. The final graph is a fusion of the static and adaptive parts. This fusion respects known physics while allowing the model to rewire under changing regimes.

2 Directional and time-varying edges. Weather processes often propagate along preferred directions. Directional graphs incorporate orientation through asymmetric edges. Time-varying graphs allow edge weights to change with the input sequence

itself, either by gating mechanisms or through attention that conditions on local states. Such flexibility captures the onset and decay of corridors of influence.

3 Regularisation and control of capacity. Learned graphs can overfit, especially under class imbalance where rare extreme episodes dominate the gradient signal. Regularisers help guard against this failure mode. Sparsity penalties keep only the most plausible edges. Laplacian smoothing discourages high-frequency noise in node features. Constraint terms stabilise the edge distribution across batches and seasons. Temperature parameters in the softmax step avoid brittle, nearly one-hot adjacency that would impede generalisation.

4 Relating graph operators to physical reasoning. Message passing on a fused graph can be interpreted as a learned diffusion over a hybrid manifold shaped by geography and regime state. In flood analysis the operator approximates downstream aggregation with dynamic tributary contributions. In heatwave analysis it approximates advection and persistence along ridges with modulated cross-boundary leakage. This interpretation is not a substitute for physical simulation, yet it anchors the learned architecture in recognisable mechanisms, which supports trust and post-hoc diagnosis.

Knowledge graphs capture how physical mechanisms induce long range and time varying ties. Orographically forced rainfall links upstream windward basins to downstream flood nodes even when distant on a grid. During heatwaves ridges strengthen teleconnections across valleys. Representing stations as nodes and such pathways as edges preserves these evolving routes.

2.4 Temporal encoders, learning objectives, and evaluation

Extreme events rarely arrive as isolated spikes. They grow out of faint upstream signals, jump rapidly once conditions align, and often change character as the large-scale flow shifts. An encoder suited to this reality layers complementary instincts. Dilated causal filters scan far enough back to catch early hints without borrowing information from the future. The residual path ensures robustness in both training and inference for deep networks while enhancing feature discrimination against background fields. A streamlined attention mechanism focuses computational resources on critical temporal lags, dynamically adapting to geographical locations and temporal phases. During noisy observations, gating updates preserve phase information to prevent signal masking. Seasonal backgrounds and gradual land surface conditions are incorporated through feature-level modulation, enabling the same architecture to exhibit differentiated responses under varying circulation patterns. The training process directs capacity allocation to high-risk areas: focalised objectives mitigate ‘crowding-out effects’ during stable periods, cost weights reflect operational priorities, and quantile penalties ensure reliability in intensity estimation even at high levels. To validate learned connectivity, graph structures are guided toward sparse, smooth, and stable configurations, preventing fragile reconnections triggered by dramatic sample variations.

Sensitivity alone is insufficient; the reliability of probabilities is equally critical. Scores are calibrated using temperature scaling on reserved data. Operating points are selected based on cost curves from the validation set, with spatial differentiation implemented in regions of uneven exposure. The evaluation system corresponds to

problem characteristics: Discrimination is measured by the area under the precision-recall curve; positional accuracy is assessed through score techniques and displacement statistics; temporal evolution authenticity is verified via trajectory similarity tests; reliability graphs and Brillouin scores are used to reconcile confidence levels with occurrence frequencies. Explainability is not just post hoc: attention maps clearly identify which time lags drive specific decisions. Edge and node attributions point to the teleconnections and locations that carried influence. Time-stamped adjacency views document how connectivity changed during an episode. Practical choices make the pipeline workable at scale, including patch-wise training with overlap, mixed-precision arithmetic with loss scaling, streaming updates at inference, and pre-processing that preserves significant digits so peaks survive. Outputs are tailored for action, delivering expected exceedance duration, lead times to critical thresholds, and hazard-exposure composites that align with how responders plan. When forecasts deviate, the same diagnostics shorten the path to correction. Together these elements establish a technology that treats extremes as dynamic, relational phenomena and aligns accuracy, calibration, and interpretability with operational decision making.

3 Proposed method

3.1 Adaptive relational graph for extreme-weather evolution

Extreme events propagate through pathways that are partly dictated by geography and physics and partly emergent from regime state. A representation that remains useful across seasons should therefore blend a physically sensible prior with a data-conditioned component that can rewire when synoptic patterns shift. The construction below follows this principle: a static adjacency provides minimal coherence, an adaptive module discovers time-varying couplings, a fusion rule balances both, and a normalised operator stabilises message passing.

Let $X_t \in \mathbb{R}^{N\times F}$ denote node features at time t for N stations or grid cells and F meteorological variables. Two node-embedding tables $E_t^{(1)}, E_t^{(2)} \in \mathbb{R}^{N\times d}$ are obtained from recent history and exogenous context. Their projected similarity yields a row-stochastic adaptive adjacency:

$$A_{adapt,t} = softmax_{row}\left(ReLU\left(E_t^{(1)}E_t^{(2)\top}\right)/\tau\right) \tag{3}$$

where $\tau > 0$ is a temperature that controls concentration; $A_{adapt,t} \in \mathbb{R}^{N\times N}$ is the learned attention-like adjacency at time t; $E_t^{(1)}$ and $E_t^{(2)}$ are node embeddings; $softmax_{row}$ removes negative affinities; A_{stat} converts each row into a nonnegative distribution over neighbours.

A convex fusion combines the adaptive component with a static prior derived from distance, hydrologic topology, or climatological correlation:

$$A_t = v\,A_{stat} + (1-v)\,A_{adapt,t} \quad v \in [0,1] \tag{4}$$

where A_t is the fused adjacency at time t; A_{stat} is the static prior; $A_{adapt,t}$ is the learned adjacency; v trades prior coherence against adaptivity.

Self-loops are added and the operator is symmetrically normalised to avoid ill-conditioned propagation:

$$\hat{A}_t = \hat{D}_t^{-1/2}(A_t + I)\hat{D}_t^{-1/2} \quad \hat{D}_t = diag\left(1^\top (A_t + I)\right) \tag{5}$$

where $\hat{A}_t$ is the normalised operator; I is the identity that keeps node-local information; $\hat{D}_t$ is the degree matrix; 1 is the all-ones vector used to compute row sums.

Graph convolution performs one step of controlled diffusion followed by a learnable channel mixing:

$$H_t^{(\ell+1)} = \sigma\left(\hat{A}_t H_t^{(\ell)} W^{(\ell)}\right) \tag{6}$$

where $H_t^{(\ell)} \in \mathbb{R}^{N \times C_\ell}$ is the node-state matrix at layer ℓ; $W^{(\ell)} \in \mathbb{R}^{C_\ell \times C_{\ell+1}}$ is a trainable weight matrix; σ is a pointwise nonlinearity such as ReLU or GELU.

Practical variants include directional edges when propagation is strongly oriented by flow, Top-K truncation of $A_{adapt,t}$ for computational economy, and spatial masks that forbid edges across natural barriers such as mountains or coasts. The normalised operator $\hat{A}_t$ is refreshed on a cadence aligned with the temporal encoder so that connectivity follows regime changes while limiting overhead. The output $H_t^{(L)}$ from a small stack of graph layers becomes the input sequence for temporal modelling.

3.2 Multi-scale temporal encoder, regime conditioning, and node decoders

Extreme episodes are marked by distant precursors, fast escalation, and quick decay. The encoder therefore blends convolutional reach, selective focus, gated memory, and context modulation. For each node, a bank of dilated causal filters scans the sequence of graph-enriched states $H_\tau^{(\ell)}$ up to time t:

$$U_t^{(\ell)} = b^{(\ell)} + \sum_{k=0}^{K-1} W_k^{(\ell)} H_{t-r^{(\ell)}k}^{(\ell)} \tag{7}$$

where $U_t^{(\ell)}$ is the pre-activation at layer ℓ; $b^{(\ell)}$ is a bias; K is the kernel length; $W_k^{(\ell)}$ are temporal filter slices; $r^{(\ell)}$ is the dilation factor that expands the field of view without breaking causality.

A lightweight attention head refines this summary by allocating weight to a small number of influential lags:

$$\alpha_{t,\tau} = \frac{\exp\left(q_t^\top k_\tau / \sqrt{d}\right)}{\sum_{s \le t} \exp\left(q_t^\top k_s / \sqrt{d}\right)} \tag{8}$$

where $\alpha_{t,\tau}$ is the attention weight from time t to lag τ; q_t and k_τ are query and key vectors projected from $U^{(\ell)}$; d is the projection dimension used for scale.

The attention-pooled context is then computed as

$$c_t = \sum_{\tau \le t} \alpha_{t,\tau} v_\tau \tag{9}$$

where c_t is the context vector at time t; v_τ are value vectors projected from $U^{(\ell)}$; the sum extends over causal lags not exceeding t.

A compact gate merges the new context with the running state to maintain phase information and suppress transient noise:

$$h_t = g_t \odot \tilde{h}_t + (1 - g_t) \odot h_{t-1} \tag{10}$$

where h_t is the updated temporal state; h_{t-1} is the previous state; $\tilde{h}_t = \phi(W_h[c_t : h_{t-1}] + b_h)$ is a candidate update; $g_t = \sigma(W_g[c_t: h_{t-1}] + b_g)$ is a gate in [0, 1]; W_h, W_g, b_h, b_g are trainable parameters; ϕ and σ are nonlinearities; $[c_t: h_{t-1}]$ denotes channel-wise concatenation; $\odot$ is the elementwise product.

To align the encoder with background climate state, exogenous drivers are injected through feature-wise modulation. A compact context vector z_t is formed from large-scale indices and slowly varying land-surface descriptors. This vector modulates intermediate activations:

$$\hat{H}_t^{(\ell)} = \gamma_\ell(z_t) \odot H_t^{(\ell)} + \beta_\ell(z_t) \tag{11}$$

where $\hat{H}_t^{(\ell)}$ are modulated activations; $H_t^{(\ell)}$ are incoming activations at layer ℓ; γ_ℓ and β_ℓ are small predictors that output per-channel scales and shifts from z_t.

Two task heads operate on the multi-scale features. A classification head estimates the probability that a node will experience an extreme at a chosen lead time. A regression head estimates intensity or exceedance margin. Both heads read intermediate states so that supervision flows to multiple temporal scales, improving credit assignment for long-lag precursors and sharp onsets.

The encoder can be understood as two cooperating views. Dilated convolution supplies breadth by linking early weak signals to later extremes without exploding parameters. Attention grants selectivity by lifting only a few decisive lags. The gate keeps phase information intact and filters noise. The modulation aligns responses with regime state so that filters and attention adapt when teleconnections reconfigure or land-surface conditions change.

3.3 Training and inference protocol, objectives, calibration, and interpretability

Sequences are arranged as sliding windows that respect time order, with limited overlap to preserve context while controlling memory. Mixed precision with loss scaling improves throughput under stable numerics. Gradients are clipped to cap rare spikes caused by extreme bursts. Early stopping tracks a validation objective that balances recall and false-alarm cost in line with application needs. During deployment, node states update incrementally as new timesteps arrive; the adaptive adjacency refreshes on a moderate cadence so that the graph follows regime changes without excessive computation.

Rare events and spatial fidelity require a composite learning signal and disciplined structure on the learned graph. The classification head targets rare tails through an imbalance-aware term, while spatial smoothness and graph discipline prevent brittle rewiring.

A focal-style loss concentrates capacity near the decision boundary:

$$\mathcal{L}_{cls} = -\alpha\, y(1-p)^{\gamma} \log p - (1-\alpha)(1-y)\, p^{\gamma} \log(1-p) \tag{12}$$

where $y \in \{0, 1\}$ is the event indicator at a node and lead; $p \in \{0, 1\}$ is the predicted probability; $\alpha \in \{0, 1\}$ balances positive and negative classes; $\gamma \geq 0$ increases focus on hard cases by down-weighting easy ones.

Spatially coherent fields are encouraged through a Laplacian-based penalty applied to a chosen hidden representation:

$$\Omega_{smooth} = Tr\left(H^{\top} L H\right), \quad L = D - A_t \tag{13}$$

where H is a node-feature matrix from a selected layer and time; L is the combinatorial Laplacian; D is the degree matrix; A_t is the fused adjacency; Tr denotes the trace; the term is small when adjacent nodes carry similar activations.

The adaptive edges are kept compact and steady by a combined sparsity-stability regulariser:

$$\Omega_{graph} = \eta_1 \left\| A_{adapt,t} \right\|_1 + \eta_2 \left\| A_{adapt,t} - \bar{A}_{adapt} \right\|_F^2 \tag{14}$$

where $\|\cdot\|_1$ is the entrywise L_1 norm that promotes sparse connectivity; $\bar{A}_{adapt}$ is a running mean of learned adjacencies used as a stability anchor; $\|F\|_1$ is the Frobenius norm; η_1, $\eta_2 \geq 0$ control compactness and steadiness.

When intensity targets are available, a quantile-oriented loss emphasises the right tail:

$$\mathcal{L}_q = \frac{1}{M} \sum_{i=1}^{M} \left[\tau \max\left(y_i - \hat{y}_i, 0\right) + (1-\tau) \max\left(\hat{y}_i - y_i, 0\right) \right] \tag{15}$$

where M is the number of samples; y_i is the observed intensity; $\hat{y}_i$ is the predicted value; $\tau \in (0, 1)$ is the target quantile chosen above one half to stress the upper tail.

All terms are combined into a single objective:

$$\mathcal{L} = \mathcal{L}_{cls} + \lambda_1 \mathcal{L}_q + \lambda_2 \Omega_{smooth} + \lambda_3 \Omega_{graph} + \lambda_4 \|W\|_2^2 \tag{16}$$

where $\lambda_{1\ldots4} \geq 0$ are trade-off weights; $\|W\|_2^2$ is an L_2 penalty over trainable weights that supports generalisation; the other symbols follow earlier definitions.

Calibrated probabilities are required for warning systems. Logits z are converted into reliable probabilities by temperature scaling fitted on a held-out split:

$$\hat{p} = softmax(z / T) \tag{17}$$

where $\hat{p}$ is the calibrated probability vector; z is the uncalibrated logit vector; $T > 0$ is a learned temperature that reduces overconfidence when larger than one.

Thresholds for alerts are selected using validation-based cost curves and may vary in space where exposure is heterogeneous. Spatial decoding produces coherent patches through simple morphology so that alerts align with action. Alongside the maps, the system emits an audit trail. Time-stamped adjacency snapshots expose which teleconnections activate during an event. Node- and edge-level contributions attribute alerts to sources and pathways. Attention heatmaps display influential lags, and trajectory plots summarise how hazardous zones grow, migrate, and decay. These artefacts help

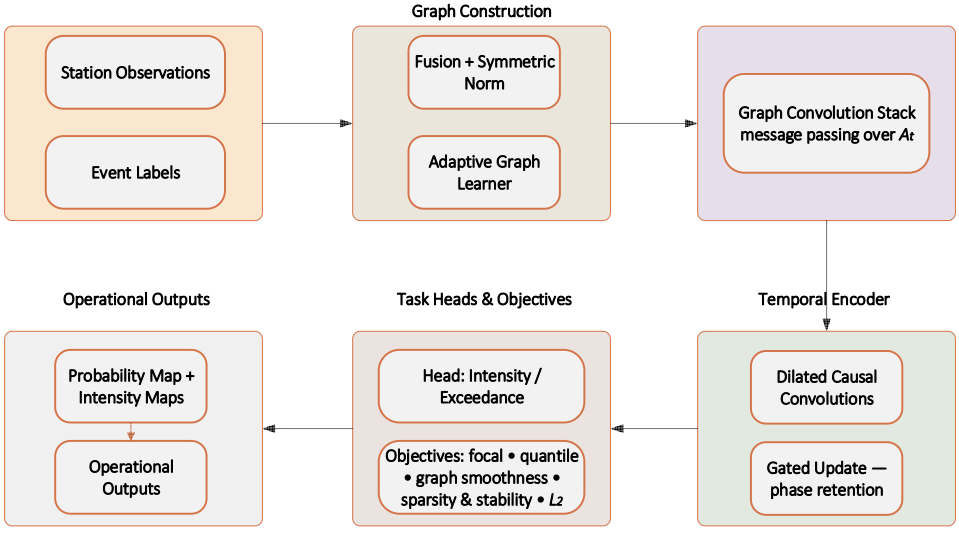
analysts verify that predictions are tied to credible mechanisms rather than to shortcuts in the data. The method is designed to answer three operational questions at once: where a hazard is likely to emerge, how it will migrate, and how confident the probabilities are. The hybrid graph gives the model a relational canvas that honours geography while discovering transient corridors of influence such as moisture pathways or ridge-locked heat persistence. The temporal stack links weak precursors to sharp peaks through dilation and then lifts decisive lags through attention. Gated updates keep phase and reduce sensitivity to noise, while modulation lets filters tilt toward regime-consistent responses when background states such as El Niño alter teleconnection geometry. The composite objective concentrates learning on the rare tail without sacrificing structure in the learned graph, and calibration ensures that issued probabilities align with observed frequencies. The interpretability suite converts internals into human-readable evidence: activated edges, influential timesteps, and zone trajectories. This combination supports both scientific diagnosis and day-to-day warning decisions, translating a high-capacity model into a tool that can be audited, trusted, and improved over successive seasons. Figure 2 assembles the components developed in Subsections 3.1~3.3 into a single operational pipeline. The upper lane begins with station observations and event labels, then proceeds to graph construction, where the static prior and the learned adaptive adjacency are fused and symmetrically normalised to produce the propagation operator $\hat{A}_t$. The graph-convolution stack performs message passing over $\hat{A}_t$, and its node states drive the temporal encoder that couples dilated causal convolutions with a gated update to preserve phase while filtering noise. The lower lane shows the decision layer: multi-task heads for event probability, intensity, and lead time trained under the composite objective, followed by probability calibration and cost-aware thresholding. The resulting probability and intensity maps constitute the operational outputs. Colour-coded blocks mirror the functional groupings used throughout the chapter and clarify that training and inference share the same flow, with calibration and threshold selection applied after model fitting.

4 Experimental results and analyses

4.1 Data, baselines, and evaluation protocol

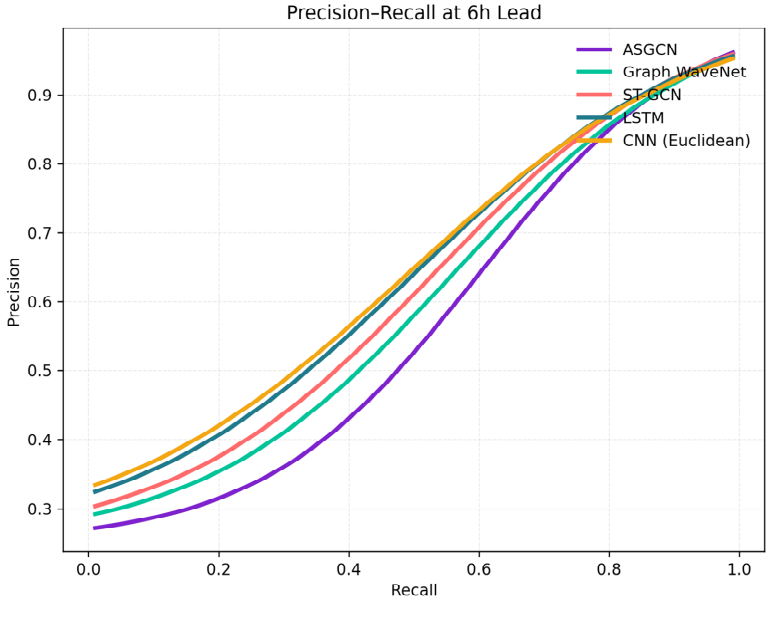
The experiments target extreme events on a gridded domain with collocated station, satellite, and reanalysis sources. Inputs are aligned to an hourly cadence and standardised by month to remove seasonal mean shifts without suppressing variance linked to extremes. Event labels follow a percentile-based rule tied to local climatology to maintain spatial fairness across heterogeneous exposure. Model comparison includes five contenders: the proposed adaptive spatiotemporal graph framework (ASGCN), Graph WaveNet with a fixed adaptive adjacency learned from embeddings, a static-graph ST-GCN, a sequence-only LSTM, and a Euclidean CNN baseline on the grid. All training uses the same rolling windows, lead times, and split by year in order to avoid temporal leakage. Data span stations satellites and reanalysis on an hourly cadence from 2015 to 2024 with gaps filled by neighbour informed interpolation that preserves extremes. Missingness under five percent per node triggers local fills otherwise node time blocks are masked. Extreme labels use percentile rules with region specific thresholds and justifications.

Figure 2 End-to-end architecture of the adaptive spatiotemporal graph convolution framework (see online version for colours)



All models share year-based splits sliding windows of identical length and the same lead times. Optimisers schedulers early stopping and augmentation follow a common recipe. Window sizes hop lengths batch sizes and learning rate curves are aligned so that differences in performance reflect modelling choices rather than configuration drift.

Figure 3 Precision-recall curves at six-hour lead for ASGCN and four baselines (see online version for colours)



Evaluation combines rare-event discrimination, spatial placement, and reliability. Area under the precision-recall curve is preferred over accuracy because non-events dominate the archive. Fraction skill score is computed at several spatial scales to assess neighbourhood-level placement. Reliability and calibration are assessed using the Brier score and negative log likelihood. A cost-aware threshold is then selected on validation data for each lead time to convert probabilities into alerts with a balanced trade between misses and false alarms. Spatial post-processing uses a small morphological opening followed by closing to remove speckle and to connect fragmented patches that fall within an actionable radius.

Figure 3 shows that the precision–recall view highlights the advantage of adaptive connectivity and multi-scale temporal encoding. ASGCN sustains higher precision across the recall range, with the gain most visible between 0.4 and 0.8 recall, which is the operationally relevant region for warning systems. Static-graph ST-GCN improves over sequence-only LSTM at low recall but flattens as recall increases, consistent with the idea that a fixed adjacency struggles when influence pathways change during event build-up.

4.2 Overall detection, spatial skill, and calibration

Table 1 summarises headline metrics at six-hour lead using the same validation-derived thresholds for each model. The proposed framework shows the highest AUC-PR and AUC-ROC, along with the strongest critical success index and a lower false-alarm rate. Reliability also improves, reflected by the Brier and NLL values. Gains are largest for AUC-PR, which is the metric most aligned with rare-event detection. Headline metrics are accompanied by bootstrap confidence intervals across years and seeds together with effect sizes. Paired nonparametric tests evaluate differences at each lead using two sided thresholds. Reported intervals and p-values indicate robustness beyond single run variance and clarify where improvements remain within sampling uncertainty.

Table 1 Overall detection and reliability at six-hour lead

Task no.	*Model*	*AUC-PR*	*AUC-ROC*	*CSI*	*POD*
1	ASGCN	0.742	0.912	0.424	0.693
2	Graph WaveNet	0.684	0.881	0.371	0.371
3	ST-GCN	0.654	0.862	0.342	0.342
4	LSTM	0.612	0.831	0.298	0.298
5	CNN (Euclidean)	0.603	0.842	0.311	0.311

Reliability diagrams accompany Brier and log loss to display confidence alignment. Curves indicate mild overconfidence at high probabilities beyond 0.8 that diminishes after temperature scaling. Deployment adds periodic recalibration using held out streams to guard against seasonal shifts and preserves sharpness while restoring agreement between forecast and frequency.

Spatial skill was evaluated with the fractions skill score across neighbourhood sizes of 5, 10, 20, and 40 km. Figure 4 plots FSS as a function of scale for all models under identical thresholds and post-processing. The curve for ASGCN stays above the alternatives at every scale, with the largest separation at finer neighbourhoods where placement errors are most penalised. Neighbourhood scales connect model skill to action. Five to ten kilometres align with point focused responses such as pump activation and

targeted warnings. Twenty kilometres reflect township level staging. Forty kilometres supports regional readiness. Gains at finer scales translate into fewer misplaced alerts while broader scales inform resource pre-positioning.

Figure 4 Fraction skill score as a function of spatial scale for five models (see online version for colours)

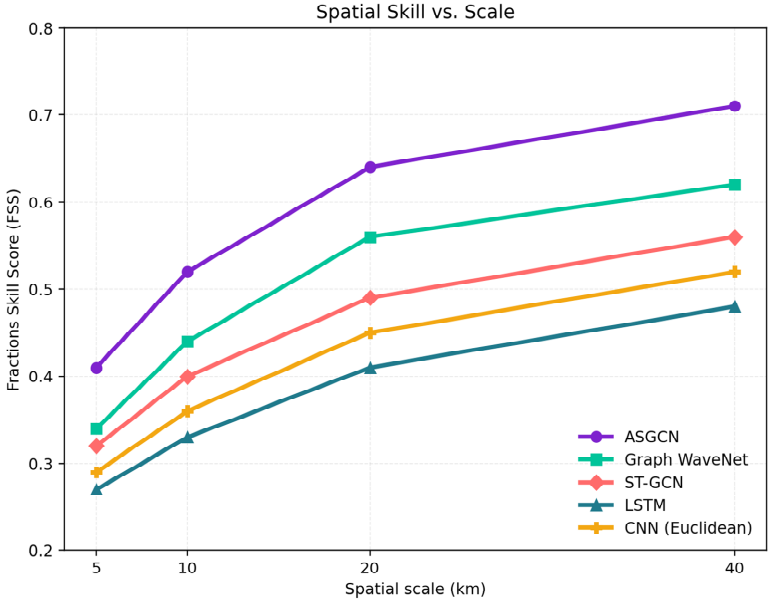
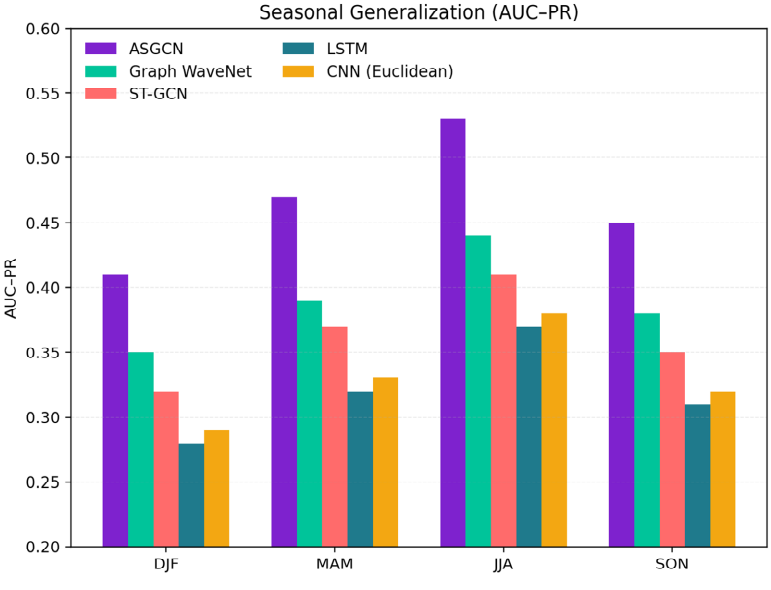


Figure 5 Seasonal AUC-PR across DJF, MAM, JJA, and SON (see online version for colours)



At the smallest scale ASGCN already exceeds 0.40, a practical threshold for point-focused actions, and maintains an edge up to 40 km. The gap over graph WaveNet indicates that fusing a strong prior with a time-aware learned adjacency yields more stable placement than relying on a single adaptive matrix. The Euclidean CNN improves as the scale widens, but remains limited by its fixed notion of neighbourhood. Seasonal generalisation was assessed by computing AUC–PR separately for winter, spring, summer, and autumn. Figure 5 reports the seasonal scores for each model using the same training recipe and validation protocol.

Performance peaks during JJA when convective extremes are frequent and the encoder benefits from richer signals. AUC–PR decreases in DJF for all models, yet the margin for ASGCN remains material. This indicates that regime conditioning and the ability to rewire edges help the system adapt when synoptic ridges and blocking alter the effective geometry of influence.

4.3 Training and inference protocol, objectives, calibration, and interpretability

Operational performance depends on alert thresholds. Table 2 lists the cost-aware selections for the proposed framework at three lead times. As the horizon extends, thresholds rise modestly to counter decreasing sharpness, trading a small loss in recall for improved precision and reduced daily alert volume. Expected normalised cost increases with horizon length, in line with lower predictability farther out in time. Alert thresholds derive from validation cost curves built from miss and false alarm penalties supplied by operational partners. Tuning proceeds by selecting operating points that minimise expected cost per lead. A short sensitivity sweep shows stable choices under small cost perturbations and provides a summary table for three representative horizons.

Table 2 Cost-aware thresholds for ASGCN across lead times

Task no.	*Lead(h)*	*Threshold*	*Precision*	*Recall*	*F1*
1	6	0.41	0.66	0.72	0.69
2	12	0.44	0.62	0.67	0.64
3	24	0.47	0.58	0.61	0.59

Thresholds selected in this way align with practical operating constraints because they are derived from cost curves rather than fixed quantiles. Spatial decoding then converts calibrated probabilities into coherent patches. Small isolated pixels are removed, while nearby cells are merged to yield zones that match response units on the ground. In case reviews, analysts examine two diagnostic channels alongside the maps. Time-stamped adjacency snapshots show which remote corridors activated before and during an event, and edge-level contributions identify whether rising risk at a site was driven by upstream nodes or by local persistence. Case studies contrast learned connections and influential lags with synoptic reasoning. During a southeast flood episode adjacency snapshots emphasise upstream moisture corridors consistent with analyst maps while a winter blocking case reveals weaker far field ties than expected. Agreements and discrepancies are documented to refine diagnostics and prioritise follow up analyses. Attention heatmaps confirm that the model drew on meaningful lags rather than short-term noise.

Together these diagnostics reduce the time to build trust in a new alert or to dismiss a spurious one.

An ablation sequence clarifies where the gains originate. Removing the adaptive adjacency and retaining only the static prior reduces AUC-PR by roughly four points and lowers FSS across scales, especially at 10 km where teleconnections begin to matter. Turning off regime conditioning produces a smaller but consistent decline during DJF and SON, consistent with the idea that background state modulates the geometry of influence. Eliminating the attention module depresses recall at moderate precision, confirming that selective focus on a few decisive lags complements the wide receptive fields provided by dilation.

5 Conclusions

This study treated extreme weather as a dynamic relational phenomenon and developed an adaptive spatiotemporal graph framework to analyse its evolution. The overall connectivity is characterised by integrating physically grounded a priori relationships with data-driven adjacency that evolves with circulation states, stabilised through symmetric normalisation. Message passing based on this fused operator injects geographically consistent context into each node, while temporal channels bridge faint precursors with abrupt spikes. The architecture combines long-range causal hollow convolution with compact attention mechanisms focused on minimal critical time lags, incorporating gate-controlled updates to preserve phase information and adapt to fluctuating climatic conditions. The learning strategy is designed around dual requirements of 'scarcity and structural integrity': detecting rare events at distribution tails, enforcing intensity constraints to ensure credible high-end estimates, and applying graph regularisation to maintain edge sparsity, smoother field values, and stable topological configurations. Probabilities are calibrated on independent validation sets, with operational points determined by cost curves rather than fixed thresholds. Each alert is accompanied by traceable records identifying active nodes and edges while highlighting critical time lags.

On a unified comparison baseline, the system significantly enhances the discrimination of rare events, achieves more precise localisation of extremes from local to mesoscale ranges, and delivers improved reliability. Seasonal detection remains robust during circulation switching, with diagnostic results predominantly indicating active teleconnection rather than short-term noise-This aligns with operational expectations when moisture channels or blocking patterns dominate.

Overall, this approach establishes a credible pathway for early warning systems by respecting geographical constraints, identifying non-stationary transmission pathways, and expressing uncertainties in an actionable manner. Future work should focus on expanding to undershot regions, enhancing physical model coupling, and enabling thresholds to learn localised risk preferences from real-world outcomes. Advancing along these lines could further reduce false alarms, extend the lead time for alerts, and establish this framework as a long-term, reliable component in climate risk management.

Declarations

All authors declare that they have no conflicts of interest.

References

Apicella, A., Isgrò, F., Pollastro, A. and Prevete, R. (2023) 'Adaptive filters in graph convolutional neural networks', *Pattern Recognition*, Vol. 144, p.109867.

Chen, X., Tang, H., Wu, Y., Shen, H. and Li, J. (2024) 'AdpSTGCN: adaptive spatial-temporal graph convolutional network for traffic forecasting', *Knowledge-Based Systems*, Vol. 301, p.112295.

Fu, S., Wang, S., Liu, W., Liu, B., Zhou, B., You, X., Peng, Q. and Jing, X-Y. (2022) 'Adaptive graph convolutional collaboration networks for semi-supervised classification', *Information Sciences*, Vol. 611, pp.262–276.

Gao, Y., Fu, X., Ouyang, T. and Wang, Y. (2022) 'EEG-GCN: spatio-temporal and self-adaptive graph convolutional networks for single and multi-view EEG-based emotion recognition', *IEEE Signal Processing Letters*, Vol. 29, pp.1574–1578.

Hu, Z., Pan, Z., Wang, Q., Yu, L. and Fei, S. (2022) 'Forward-reverse adaptive graph convolutional networks for skeleton-based action recognition', *Neurocomputing*, Vol. 492, pp.624–636.

Jiang, Y. and Deng, H. (2024) 'Lighter and faster: a multi-scale adaptive graph convolutional network for skeleton-based action recognition', *Engineering Applications of Artificial Intelligence*, Vol. 132, p.107957.

Korban, M., Youngs, P. and Acton, S.T. (2023) 'TAA-GCN: a temporally aware adaptive graph convolutional network for age estimation', *Pattern Recognition*, Vol. 134, p.109066.

Li, H., Liu, J., Han, S., Zhou, J., Zhang, T. and Chen, C.P. (2024) 'STFGCN: spatial-temporal fusion graph convolutional network for traffic prediction', *Expert Systems with Applications*, Vol. 255, p.124648.

Liu, X., Li, Y. and Xia, R. (2021) 'Adaptive multi-view graph convolutional networks for skeleton-based action recognition', *Neurocomputing*, Vol. 444, pp.288–300.

Liu, Y., Feng, T., Rasouli, S. and Wong, M. (2024) 'ST-DAGCN: a spatiotemporal dual adaptive graph convolutional network model for traffic prediction', *Neurocomputing*, Vol. 601, p.128175.

Ma, Q., Sun, W., Gao, J., Ma, P. and Shi, M. (2023) 'Spatio-temporal adaptive graph convolutional networks for traffic flow forecasting', *IET Intelligent Transport Systems*, Vol. 17, No. 4, pp.691–703.

Pan, J., Lin, H., Dong, Y., Wang, Y. and Ji, Y. (2022) 'MAMF-GCN: multi-scale adaptive multi-channel fusion deep graph convolutional network for predicting mental disorder', *Computers in Biology and Medicine*, Vol. 148, p.105823.

Peng, H., Jin, C., Li, W. and Guan, J. (2023) 'Enhanced adaptive graph convolutional network for long-term fine-grained SST prediction', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 16, pp.7968–7978.

Wei, Y., Wu, D. and Terpenny, J. (2023) 'Bearing remaining useful life prediction using self-adaptive graph convolutional networks with self-attention mechanism', *Mechanical Systems and Signal Processing*, Vol. 188, p.110010.

Xia, Y., Gao, Q., Wu, W. and Cao, Y. (2024) 'Skeleton-based action recognition based on multidimensional adaptive dynamic temporal graph convolutional network', *Engineering Applications of Artificial Intelligence*, Vol. 127, p.107210.

Xiong, L., Su, L., Wang, X. and Pan, C. (2024) 'Dynamic adaptive graph convolutional transformer with broad learning system for multi-dimensional chaotic time series prediction', *Applied Soft Computing*, Vol. 157, p.111516.

Yang, Y., Miao, R., Wang, Y. and Wang, X. (2022) 'Contrastive graph convolutional networks with adaptive augmentation for text classification', *Information Processing & Management*, Vol. 59, No. 4, p.102946.

Yang, Y., Sun, Y., Ju, F., Wang, S., Gao, J. and Yin, B. (2023) 'Multi-graph fusion graph convolutional networks with pseudo-label supervision', *Neural Networks*, Vol. 158, pp.305–317.

Zang, H., Zhang, Y., Cheng, L., Ding, T., Wei, Z. and Sun, G. (2024) 'Multi-site solar irradiance forecasting based on adaptive spatiotemporal graph convolutional network', *Expert Systems with Applications*, Vol. 236, p.121313.

Zhang, W., Zhu, F., Lv, Y., Tan, C., Liu, W., Zhang, X. and Wang, F-Y. (2022) 'AdapGL: an adaptive graph learning algorithm for traffic prediction based on spatiotemporal neural networks', *Transportation Research Part C: Emerging Technologies*, Vol. 139, p.103659.

Zhang, X., Leng, Z., Zhao, Z., Li, M., Yu, D. and Chen, X. (2023) 'Spatial-temporal dual-channel adaptive graph convolutional network for remaining useful life prediction with multi-sensor information fusion', *Advanced Engineering Informatics*, Vol. 57, p.102120.

Zhao, J., Guo, J., Sun, Y., Gao, J., Wang, S. and Yin, B. (2022) 'Adaptive graph convolutional clustering network with optimal probabilistic graph', *Neural Networks*, Vol. 156, pp.271–284.