Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Multimodal learning with lifecycle awareness for early misinformation detection in social information systems

Di Wu

wudi19450163.com

Officers College, Chinese People's Armed Police Force

Xiaoyu Su

Officers College, Chinese People's Armed Police Force

Ziyang Wang

Officers College, Chinese People's Armed Police Force

Research Article

Keywords: multimodal deep learning, misinformation detection, lifecycle modeling, contrastive learning, early warning, social information systems, event-state management

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11146972/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Multimodal learning with lifecycle awareness for early misinformation detection in social information systems

Di Wu[1,*], Xiaoyu Su[1], Ziyang Wang[1]

[1]Officers College, Chinese People's Armed Police Force, Chengdu, 610213,China

*Corresponding author: Di Wu, wudi1945@163.com

Abstract: Misinformation often causes harm before rich propagation evidence becomes available, making early detection difficult for content-governance systems. This study proposes CIML-Net, a lifecycle-aware multimodal framework for early misinformation detection. The method divides an event into weak lifecycle stages using time, cumulative size, and growth-rate statistics, then encodes textual, visual, and event-structure signals with RoBERTa, CLIP, and a graph attention network. Stage-conditioned affine fusion and cross-stage contrastive learning learn representations that remain stable as an event evolves. Experiments use event-disjoint evaluation on Weibo21 and PHEME, with CHEF serving as text-only transfer validation, and compare CIML-Net with propagation-only, multimodal, and graph-contrastive baselines as well as three recent journal methods published in 2025-2026. CIML-Net improves early-window detection over all of them: on Weibo21 it achieves 80.1% Early-F1@6h and 91.0% full-trajectory F1, exceeding the strongest recent baseline by 2.4 points in the 6 h window, and mechanism-matched ablations link the early-window gain to the cross-stage positive definition under fixed encoder capacity. Event-cluster bootstrap intervals exclude zero for the principal early-window deltas, and a nine-fold leave-one-macro-event-out evaluation on PHEME preserves the method's aggregate advantage under unseen macro events. Results indicate that lifecycle-conditioned multimodal learning improves early-window detection under the evaluated benchmarks and a bounded single-node serving setting.

Keywords: multimodal deep learning, misinformation detection, lifecycle modeling, contrastive learning, early warning, social information systems, event-state management

1. Introduction

1.1. System-level challenges in content governance

Large social platforms ingest user-generated content as a continuous stream, and recent reviews frame misinformation detection as a problem spanning detection, verification, mitigation, and cross-modal evidence integration [1,2]. When a story breaks during a public-health emergency, a financial rumor, or a politically sensitive incident, a risk-detection service has to encode each incoming post, keep the event-level state current, judge whether the event is still emerging or already spreading, and return a moderation signal inside a fixed latency budget. Classifier accuracy is only one part of the operational requirement. In the serving setting considered here, the system must also sustain low-latency inference under bursty traffic, maintain a consistent event-level state, and remain usable when only limited propagation evidence has accumulated.

Many misinformation benchmarks are evaluated after substantial propagation evidence has accumulated. Such post-hoc settings favor methods that can use complete repost graphs, mature diffusion

statistics, or offline batch computation. Early-detection work instead evaluates truncated evidence before a cascade matures [3]; high full-trajectory F1 therefore does not by itself establish early-deployment utility. CIML-Net is designed for this mismatch. It is not proposed as a distributed stream-processing architecture; instead, the detector is evaluated together with a bounded serving study on real benchmark event sequences, so that predictive performance and model-side system requirements are reported under the same early-event conditions.

1.2. Early intervention and lifecycle structure

Timing is a central constraint because false news can diffuse rapidly and broadly [4], while health misinformation can create direct public-health risks [5]. AI-based digital disinformation also poses a growing information-systems challenge [6]. Recent work on generative language models further highlights how fluent, difficult-to-verify misinformation can be produced at lower cost and larger scale [7]. Early intervention is therefore required under the least informative conditions: the propagation graph is still sparse, and the available evidence may consist of a short post or a single image. Paraphrases, edited images, fact-check replies, and large repost cascades typically appear later.

We treat that temporal change as a sequence of observable event states rather than as a single static sample. Although surface evidence changes as an event develops, the central claim and much of its persuasive framing can persist. Recurrent rumor models have already shown that incomplete microblog trajectories contain useful information before a cascade is complete [8]. CIML-Net uses a different training relation: slices drawn from different lifecycle stages of the same event form positive pairs, encouraging the representation to remain tied to event identity as the available evidence changes.

1.3. Contributions and scope

The paper makes three contributions, each deliberately narrow. The first is a way to operationalize misinformation propagation: a five-stage weak lifecycle, comprising incubation, initiation, eruption, sustainment, and decay, derived from time, cumulative event size, and growth-rate statistics. We do not present these as social stages we discovered; they are an engineered signal, and the numerical cut points are heuristic operating points introduced for reproducible weak conditioning rather than universal diffusion constants. The dataset evidence boundaries used in the experiments are specified in Section 5.1. The second is the model itself: cross-stage positive construction together with stage-conditioned affine fusion for early-window multimodal detection. Earlier early-detection methods modeled propagation paths or learned credible cues from incomplete cascades [9,10]; our contribution is to couple lifecycle-conditioned multimodal fusion with same-event, cross-stage positives under event-disjoint evaluation. The combination of contrastive learning, multimodal encoding, and propagation-graph modeling is incremental, but the same-event, different-stage positive-pair definition is, to our knowledge, new and, more importantly, testable. The third contribution is the set of controls that make the lifecycle claim auditable rather than merely asserted: event-disjoint splits; ζ removal; random-stage labels; threshold perturbation; graph sensitivity; a content-free propagation-statistics baseline (Propagation-only LR/GBDT); protocol-matched re-runs of the graph-contrastive GACL and of three recent journal baselines published in 2025-2026 that cover balanced multimodal interaction (BCMPL), temporal-spatial contrastive learning (TSHCL), and propagation-prediction-based early detection (PreERD); mechanism-matched ablations that separate cross-stage positives from graph-augmented, cross-modal, and random

alternatives (Section 5.6); and a per-sample prediction-consistency analysis that asks whether the two models are complementary or just reusing the same shortcut (Section 5.9). Per-seed breakdowns appear in Section 5.8, and the system measurements are recorded on a single-node deployment over real benchmark event sequences and reported separately from the algorithmic evidence.

2. Related work

2.1. Multimodal and propagation-based misinformation detection

Multimodal misinformation detection pairs textual and visual signals. Matching-aware co-attention and mutual distillation improve cross-modal alignment [11], while modality-neutralization-aware fusion addresses feature dilution [12]. Knowledge-aware pre-training and balanced cross-modal prompt learning further strengthen semantic interaction [13,14]. Recent work also enhances large-language-model-based fake-news video detection through cross-modal evidence retrieval [15]. These methods primarily strengthen content-level representation, but they generally do not model how one misinformation narrative changes as it moves from seed content to mass diffusion.

A second line of work reads the propagation itself, through retweet, reply, and conversation trees. Bi-GCN remains a representative graph baseline for top-down and bottom-up structure [16], and graph-adversarial contrastive learning has added contrastive objectives for propagation robustness [17]. Recent work goes further by predicting the not-yet-formed propagation structure to compensate for the missing early graph [18]. These methods work well once a rich enough graph has accumulated. Their weak spot is the first few hours, when the graph barely exists, and on cross-platform or news-based data the original repost chains may never be available at all, which is exactly the regime an early-warning system lives in.

2.2. Contrastive and temporal representation learning

Representation learning for rumors increasingly exploits temporal or structural relations rather than only class labels. Recurrent models encode sequential change across a developing microblog event [8], while graph-adversarial, temporal-spatial, and cross-domain contrastive methods define positives through graph views, behavioral neighborhoods, or domain adaptation [17,19,20]. CIML-Net defines positive pairs using event identity across observed lifecycle stages: two slices from the same event are paired even when their stage-specific evidence differs. The intended invariance is therefore anchored to the event's own temporal evolution rather than to a class label, a generic augmentation, or an unrelated video-level view.

2.3. Early and online misinformation detection

Early-detection research evaluates what can be inferred before a cascade matures. Propagation-path models classify short user sequences [9], whereas credible early-detection methods seek reliable decision points from incomplete repost sequences [10]. Adversarial information-campaign generation augments sparse or conflicting early responses [21]. Historical-pattern retrieval transfers signals from past rumors to emerging events [22]. These studies make temporal truncation a first-class concern, but they do not jointly model lifecycle-conditioned multimodal representations and deployment feasibility. CIML-Net targets that narrower gap; the serving analysis is presented as a bounded single-node evaluation over real benchmark event sequences, not as a contribution to general stream-processing or LLM-serving systems. Recent studies have further strengthened both multimodal representation learning and early propagation modeling. BCMPL employs balanced cross-modal prompt learning to improve task-specific text-image

interaction while mitigating modality imbalance [14], whereas TSHCL exploits hierarchical temporal-spatial contrastive signals from public behavior [19]. More recently, PreERD predicts partially unobserved propagation structure before rumor classification, explicitly targeting the sparse-evidence regime of early detection [18]. These methods represent three closely related directions, namely multimodal interaction, temporal-spatial contrastive learning, and early propagation prediction. We therefore include protocol-adapted implementations of these recent methods in the experimental comparison, while using mechanism-matched ablations to distinguish their underlying mechanisms from CIML-Net's lifecycle-conditioned cross-stage learning.

Two very recent Weibo21 methods deserve explicit mention. GC2MFND [23] combines multi-granularity conflict modeling with domain-guided calibration and reports a 0.957 F1 on Weibo21, and MVFA [24] reports a 0.9326 F1 with multi-view fusion attention over text-derived views. Their published evaluations do not use an event-disjoint, early-window protocol comparable to ours, so those values are not directly comparable with the protocol-matched results presented later. They are therefore treated only as non-protocol-matched references in the property-level comparison of Section 7.1. Protocol-matched re-implementations of these methods under event-disjoint, early-window conditions are an important next step but are beyond the scope of this revision.

3. Lifecycle formulation

3.1. Operational definition

We treat a misinformation event as a time-ordered collection of posts, news items, replies, or event slices that share a common claim. So that the staging stays reproducible across public datasets, we derive lifecycle stages from only three quantities: time offset, cumulative observed node count, and growth rate. Table 1 lists the five stages; the weak-label rules are applied as an ordered decision list (Section 3.1), so that every observed (time, size, growth) tuple maps to exactly one stage. Nodes correspond to observed platform items in both benchmarks: Twitter posts or replies in PHEME [25], and source samples from the Weibo21 benchmark [26]. For Weibo21, the repost/comment relations used to form the event graph were separately re-collected by the authors, as detailed in Section 5.1.

Table 1. Lifecycle stages and weak-label rules.

Stage	Operational meaning	Observable signals	Weak rule
Incubation	Seed content; few participants	content and initial timestamp	R5: else (no N condition); reached only when $\Delta\tau \leq 2$ h
Initiation	Concentrated reposting or related items begin	increasing event count	R4: $\Delta\tau > 2$ h and $N_{posts} \leq 100$ (if R1–R3 not met)
Eruption	Large-scale expansion	rapid growth in event graph	R2: $\Delta\tau \leq 48$ h and $N_{posts} > 100$ (if R1 not met)
Sustainment	Continued spread with slower growth	declining hourly additions	R3: 48 h $< \Delta\tau \leq 168$ h (if R1–R2 not met)
Decay	Event becomes inactive	near-zero new nodes	R1: $\Delta\tau > 168$ h, or growth < 10% of prefix peak

Note. Rules are evaluated as an ordered decision list from R1 to R5; the first matching rule assigns the stage, with R5 as the else-clause. The numerical cut points are heuristic operating points; their rationale and robustness checks are described in the text below.

The numerical cut points in Table 1 are heuristic operating points introduced in this study rather than empirical constants imported from a prior social-stage taxonomy. The 2 h boundary isolates the immediate seed period, 48 h defines a coarse two-day expansion window, 168 h sets a one-week tail horizon, Nposts = 100 separates small from already large observed cascades, and the 10% prefix-peak rule provides a scale-relative inactivity criterion. These values are used only to generate auxiliary weak conditioning labels and are not interpreted as universal diffusion thresholds. The 12 h and 24 h values used later are evaluation windows, not lifecycle boundaries. Sensitivity to the time/size cut points is examined later in Sections 5.7 and 5.8 using +/-10% and +/-20% perturbations, so the reported result is not tied to a single exact time/size setting.

The five rules in Table 1 are evaluated as an ordered decision list, R1 through R5, and the first matching rule assigns the stage. Because the list ends in an else-clause, every legal ($\Delta\tau$, N, growth) tuple maps to exactly one stage: a 20-h slice with 50 posts is Initiation via R4, and a 6-h slice that already has 150 posts is Eruption via R2. Assignment is per slice and carries no monotonicity constraint: an event whose hourly growth falls below 10% of the prefix peak enters Decay via R1, and a later growth recovery can return subsequent slices to an earlier stage. In the Weibo21 event inventory, 6.8% of multi-slice events exhibit at least one such backward transition; we retain this behavior deliberately, because a recovering event genuinely re-enters an earlier evidence regime, and the stage labels act as weak conditioning signals rather than a formal monotone state machine. Rule order also subsumes the growth bands, since R1 has already removed any slice growing at less than 10% of the prefix peak. Near a boundary the hard rule is relaxed: a slice sitting just under the 2 h mark, for instance, gets shared probability mass over incubation and initiation rather than a single stage assignment. This boundary smoothing reduces discontinuities introduced by hard cutoffs and better reflects continuous diffusion. The resulting labels are auxiliary weak labels, not empirically discovered social stages; their role is only to provide reproducible conditioning information for the model. Figure 1 shows both the five-stage scheme and the weak-label procedure.

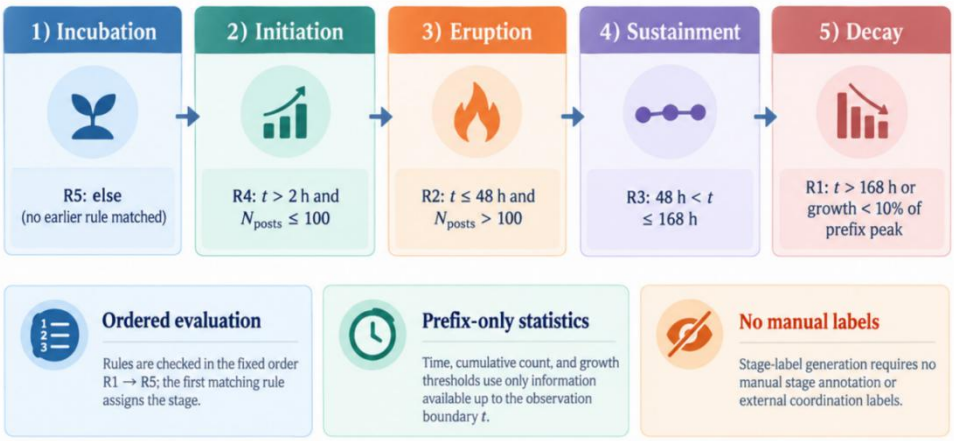


Figure 1. Five-stage lifecycle and weak-label generation.

3.2. Formal event representation and stage assignment

Each event is formalized as $e = (P, G, T)$, with P denoting the observed items, G the event-structure graph, and T the timestamp sequence. Across datasets, edges in G are observed propagation relations such as reposts, replies, or conversation links. At time t, the detector receives only $P_{\le t}$, $G_{\le t}$, and $T_{\le t}$. The classifier $f(P_{\le t}, G_{\le t}, T_{\le t}) \rightarrow \{real, fake\}$ is therefore evaluated from partial event states, including the 6 h window after first observation.

Stage assignment uses three observable quantities. The time offset is $\Delta\tau = t - \tau_{first}$, cumulative size is Nposts = |{pi : τi ≤ t}|, and the third input is hourly growth. Wherever the stage rules refer to peak growth, the peak is the prefix peak g_peak(t) = max{g(u) : u ≤ t}, i.e., the maximum hourly growth observed up to the current time t; the full-trajectory maximum is never used, so the sustainment and decay rules in Table 1 introduce no future information. Together they determine a hard stage, via the first matching rule in Table 1, and a boundary-smoothed soft label. Because time and size may also proxy propagation speed or cascade scale, these variables create a potential shortcut. Section 5.2 therefore includes a propagation-statistics baseline that uses no text, image, or graph encoder and directly measures how much performance can be obtained from those statistics alone.

4. CIML-Net

4.1. Architecture overview

CIML-Net is organized as four modules: a multimodal content encoder, a propagation/event-graph encoder, a stage-inference-and-conditioning module, and a multi-task output head; Figure 2 traces the data flow between them. Given one event slice, the model emits both a misinformation probability and an auxiliary stage distribution. Cross-stage contrastive learning runs at training time only. Online, the system reads the current event state and nothing from the future.

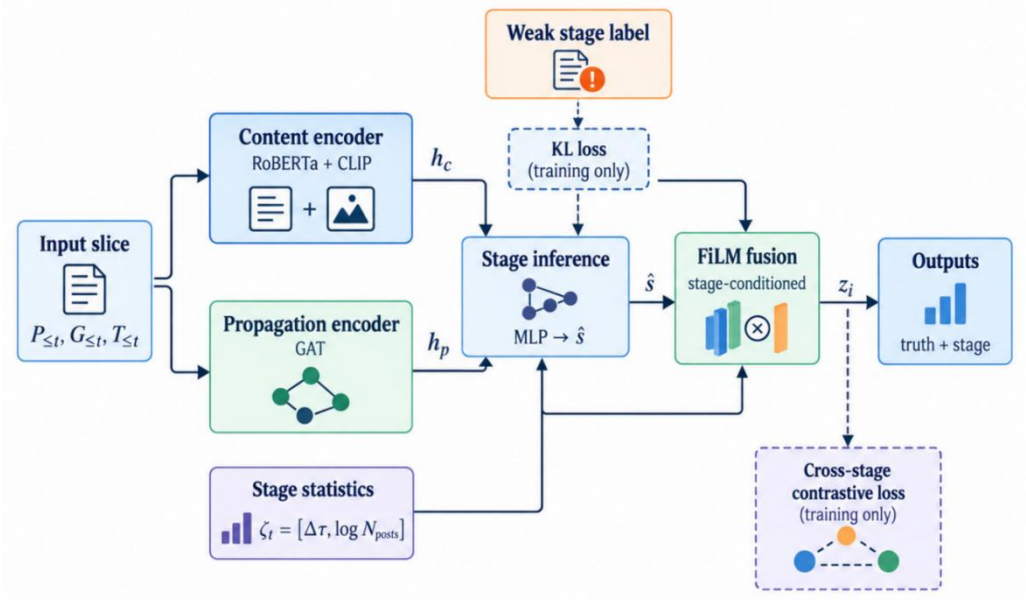


Figure 2. CIML-Net architecture. Cross-stage contrastive learning is a training-only objective.

4.2. Multimodal content encoding

Each observed item pi is encoded with a BERT-family pretrained transformer [27]: RoBERTa is used for English data [28], while Chinese-RoBERTa-wwm-ext is used for Chinese data [29]. Recent fake-news work also supports RoBERTa under cross-dataset generalization settings [30]; the [CLS] vector is

taken as the text representation. If the item carries an image, CLIP ViT-B/16 produces the visual vector, following recent CLIP-based multimodal fake-news detection frameworks [31]. The CLIP visual encoder is kept frozen in all CIML-Net runs, while the RoBERTa text encoder and the downstream fusion, graph, stage, and classification modules are optimized on the training split. If the item does not carry an image, we substitute a learnable missing-image embedding rather than zeroing the slot. Bidirectional cross-attention then lets text attend over the image features and the image attend back over text, and an MLP over both directions yields the fused content vector mi. These image features are most relevant at the earliest stage, when the text may be short or ambiguous. In such cases, a visual cue—for example, a forged document, a doctored screenshot, or a misleading image—can provide information that is absent from the text.

4.3. Propagation-graph encoding

The graph encoder reads event structure. For PHEME, edges follow the released conversation relations [25]. For Weibo21, the base sample IDs come from the official benchmark [26], while the repost/comment links used here were re-collected by the authors from the Weibo API (Section 5.1). A two-layer GAT with eight attention heads encodes each node against its neighbors, consistent with recent graph-attention approaches to fake-news detection [32]. Early on, though, the graph is barely there. When the observation window is short and the structure collapses to a handful of isolated nodes, the graph branch simply falls back to a transformed node feature instead. This fallback is deliberate: a detector that fails as soon as diffusion evidence is missing is of no use in the first few hours, which is the regime we care about.

4.4. Stage-conditioned affine fusion

From the content vector, the graph vector, and an optional temporal-size feature $\zeta i = [\Delta \tau i, \log N\text{posts}]$, the stage module estimates a distribution si over the five lifecycle stages; during training we pull si toward the weak stage label with a KL term. That stage distribution then steers the final fusion through a learned feature-wise affine transform, with $\gamma(si)$ and $\beta(si)$ rescaling and shifting the concatenated content and graph features. The effect is that the model can lean on content during incubation and initiation, when the propagation graph carries little signal, and shift its weight onto propagation structure during eruption and sustainment, when the structure is informative. Since ζi comes from the same statistics that generate the stage labels, we do not treat it as a free feature; instead, we probe its contribution: the ζ-removed and ζ-only variants let us estimate how much of the performance is lifecycle-conditioned multimodal representation and how much is just a direct read on time and size. We do not report stage-classification accuracy as a separate result, since the labels it would be scored against are themselves rule-generated.

4.5. Cross-stage contrastive objective

Within a mini-batch we form a positive pair whenever two samples come from the same event at different lifecycle stages; samples from any other event serve as negatives. The point is to push the representation onto the narrative core that an event keeps, not the surface statistics of whichever stage it happened to be sampled at. Write z_i for the normalized final representation of anchor i, $P(i)$ for the same-event, different-stage positives, and $A(i)$ for every other sample in the batch except i. The supervised contrastive objective is then:

$$L_{con} = -\sum_{i} \frac{1}{|P_{(i)}|} \sum_{p \in P_{(i)}} log \frac{exp(z_i{}^T z_p/\tau)}{\sum_{a \in A_{(i)}} exp(z_i{}^T z_a/\tau)} \quad (1)$$

The denominator follows the temperature-scaled contrastive form used in graph-contrastive rumor learning [17]; what changes is the positive mask, set by same-event, different-stage membership. When an event lands only one stage in a batch, P(i) is empty and that anchor drops out of the contrastive term. Stage-balanced sampling keeps this rare: for most events a batch holds at least two stages, and no single stage is allowed to dominate.

$$L = L_{cls} + \alpha L_{stage} + \gamma L_{con} \quad (2)$$

where $\alpha = 0.5$ and $\gamma = 1.0$ are selected on the validation set. Figure 3 illustrates the cross-stage positive and negative construction.

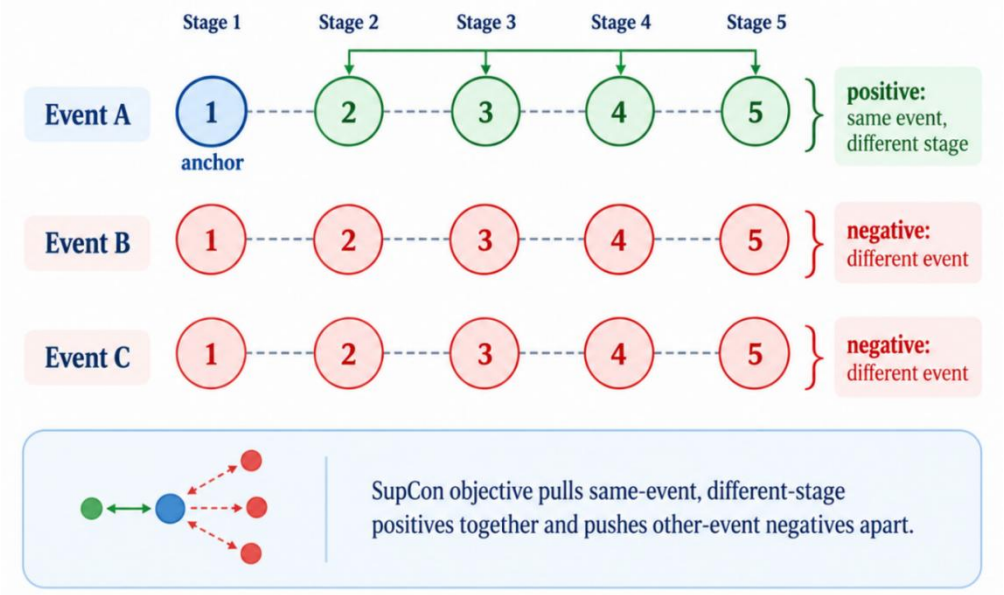


Figure 3. Cross-stage positive and negative construction.

5. Experiments

5.1. Datasets and interpretation boundaries

We evaluate on three secondary benchmark datasets, each serving a different role, but their access conditions and evidence boundaries must be stated precisely. PHEME and CHEF are publicly accessible secondary datasets, whereas access to the original Weibo21 benchmark is provided by its maintainers through an application process. In addition to these benchmark data, the Weibo21 experiments use author-recollected propagation metadata obtained through the Weibo API for sample IDs contained in the Weibo21 release. These repost/comment cascade metadata are not part of the official Weibo21 release. Accordingly, this study combines secondary benchmark datasets with author-recollected propagation metadata; no participants were directly recruited or contacted as part of the study. Weibo21 is the main Chinese multimodal benchmark in this study [26]. The official release provides text, images, labels, and timestamps, but not timestamp-level repost/comment cascades. For lifecycle-stage construction, we therefore use a locally processed real_cascade version built by supplementing the benchmark with the author-recollected Weibo API metadata described above. The corpus contains 4,488

fake and 4,640 real news samples (9,128 in total); the reconstructed cascade version comprises 690 retained label-pure event clusters with representative images, timestamps, and repost/comment cascades. The cascades are reconstructed from author-recollected Weibo API metadata as follows: (1) repost chains are extracted from the retweet_count, repost_timeline, and user interaction fields in the collected raw JSON responses; (2) timestamps are parsed from the created_at field and converted to UTC+8; (3) events are clustered by temporal proximity (posts within 48 h sharing cosine similarity > 0.75 in the Chinese-RoBERTa embedding space) using HDBSCAN (min_cluster_size=5); clustering uses no veracity labels, but clusters found after the fact to mix fake and real samples are discarded, so every retained cluster is label-pure by construction (purity, defined as the maximum within-cluster class share, equals 1.0); (4) isolated posts not assigned to any cluster are treated as single-post events and retained only if they have a representative image. Under the main configuration this pipeline yields 764 candidate clusters, of which 74 label-mixed clusters (512 member samples) are discarded, so the retained evaluation inventory comprises 690 label-pure clusters covering 8,616 of the 9,128 released samples (full audit in Appendix B). All Weibo21 experiments reported below are computed on this retained 8,616-sample inventory rather than the raw 9,128-sample release. Because the raw API responses were re-collected by the authors and are not distributed with Weibo21, holders of the official release cannot regenerate identical cascades on their own; we will make available upon acceptance the reconstruction script, the configuration file, the list of sample IDs for which cascades were recovered, and the recovered cascade metadata, so that the exact event inventory used in this paper can be audited and reused. PHEME is used as an English propagation benchmark. We use the nine-event PHEME Rumor Detection and Veracity Classification release [25] rather than the older five-event release. The nine-event release contains 6,425 conversation threads, including 2,402 rumor and 4,023 non-rumor threads, and approximately 105,354 tweets when source and reaction tweets are counted. These threads belong to only nine breaking-news macro events, the granularity at which the official release is organized; the implications for split design are addressed in Section 5.2. CHEF is an evidence-based Chinese fact-checking dataset with 10,000 claims and evidence; it has no propagation graph or image modality and is used only as text-oriented transfer validation [33]. Table 2 summarizes the three datasets, their evidence boundaries, and the role each plays.

Table 2. Dataset roles and evidence boundaries.

Dataset	Language	Events/claims/samples	Visual evidence	Graph evidence	Role
Weibo21	Chinese	raw release: 9,128 samples; retained evaluation inventory: 8,616 samples in 690 label-pure clusters	complete representative images	author-recollected repost/comment cascades	main multimodal benchmark
PHEME	English	9 events; 6,425 threads; ~105k tweets; 2,402 rumor / 4,023 non-rumor	limited / partial images	real Twitter conversation trees	propagation generalization; corrected stats
CHEF	Chinese	10,000 claims with evidence	none	none	text transfer validation only

5.2. Evaluation protocol and leakage control

Our headline aggregate metric is full-trajectory F1, with Early-F1@6h, Early-F1@12h, and Early-F1@24h reporting the early window. Every split is event-disjoint, so no event cluster or conversation thread is allowed to span training, validation, and testing, at a 7:1:2 ratio repeated over seeds 42, 123, 456, 789, and 2024. The split unit is the retained label-pure event cluster on Weibo21 and the

conversation thread on PHEME; each seed therefore partitions the 690 clusters (8,616 samples) into approximately 483 training, 69 validation, and 138 test clusters, with per-seed sample counts varying slightly with cluster sizes and released together with the split files. Because PHEME's 6,425 threads belong to only nine breaking-news macro events, thread-level disjointness does not by itself establish generalization to unseen macro events; we therefore complement the thread-level split with a grouped leave-one-macro-event-out evaluation reported later in Section 7, whose results temper but do not overturn the propagation-pattern evidence. Baselines see the identical event split and the identical early-window inputs. At training time a historical event can contribute several lifecycle stages to the contrastive loss; at test time a fresh event is truncated to the chosen observation window. Because the stage labels are built partly from time and cumulative size, the protocol has to carry a zero-content propagation-statistics baseline. It is not a CIML-Net variant at all: it is a standalone logistic-regression and gradient-boosted-tree (GBDT) model trained on nothing but $\Delta\tau$, log Nposts, hourly growth, cumulative count, and a few simple early-cascade statistics. Should this baseline come close to CIML-Net's Early-F1@6h, the lifecycle claim would have to be downgraded from multimodal narrative learning to plain propagation-speed exploitation.

Means and standard deviations sit next to the per-seed breakdowns in Section 5.8, so a reader can gauge run-to-run stability and confirm that the trends are not the product of one lucky seed. The latency, FLOPs, QPS, and sharding figures are measured on a single-node A100 deployment driven by real benchmark event sequences, with the warm-up and batch policies, queueing statistics, and full latency distributions recorded in the raw run logs.

A compact overview of this protocol is shown in Figure 4, which summarizes the Weibo21 cascade reconstruction, label-purity audit, retained evaluation inventory, and leakage-controlled split/evaluation setup together with the PHEME thread-disjoint protocol.

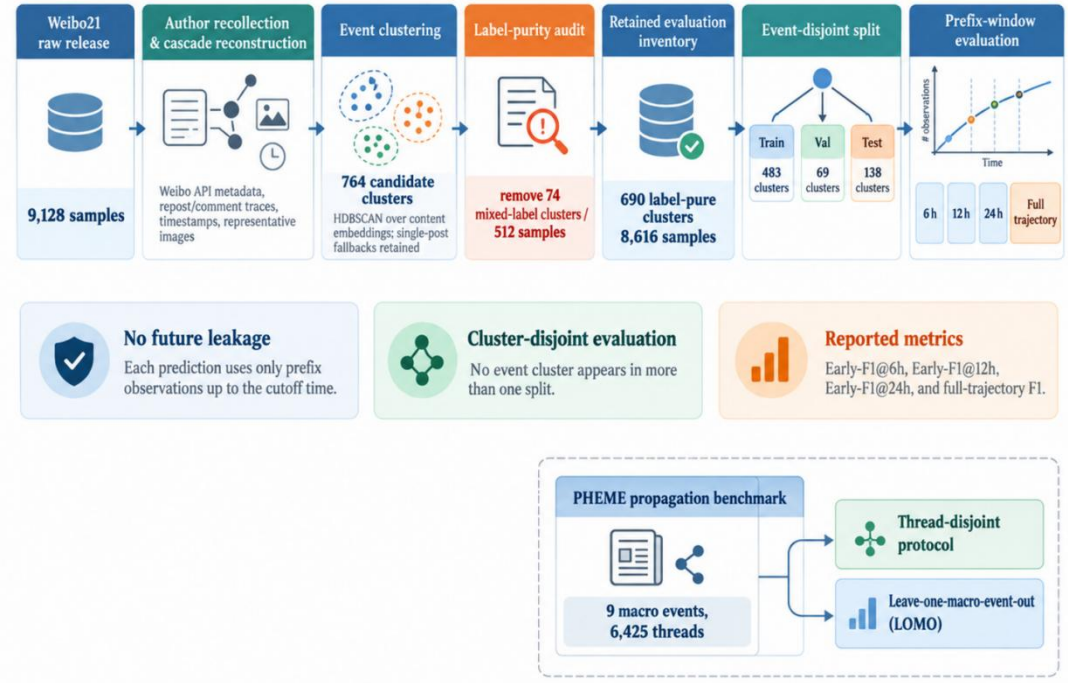


Figure 4. Data construction and leakage-controlled evaluation protocol.

5.3. Baselines and implementation

The baseline suite is organized into classical, recent, and diagnostic comparison groups. BERT [27] and RoBERTa [28] serve as the content-only transformer references; Chinese-RoBERTa-wwm-ext [29] is used for the Chinese encoder configuration. SpotFake [34] and CAFE [35] cover established multimodal fusion; Bi-GCN represents propagation-graph modeling [16]; BMR provides a multi-view multimodal baseline [36]; and GACL is the graph-contrastive comparator [17]. Propagation-only LR and GBDT are retained as diagnostic controls that estimate how much early predictive performance can be obtained from cascade statistics without content encoders.

To strengthen comparison with recent methods, we additionally include three journal algorithms published in 2025-2026. BCMPL represents recent balanced cross-modal prompt learning for multimodal fake-news detection [14]; TSHCL represents temporal-spatial hierarchical contrastive learning from public propagation behavior [19]; and PreERD represents propagation-prediction-based early rumor detection [18]. These methods are selected because they correspond directly to the three main sources of signal considered by CIML-Net: multimodal content, contrastive temporal-structural representation, and incomplete early propagation.

For fair comparison, all reproducible baselines are evaluated using the same event-disjoint data split and the same observation windows as CIML-Net. The original learning mechanism of each recent method is retained wherever possible, while dataset adapters are introduced only to accommodate the available modalities and event representation. No future observations beyond the corresponding 6 h, 12 h, or 24 h cutoff are supplied at test time. For a method whose required modality is unavailable on a dataset, the corresponding result is marked as unavailable rather than imputed. Appendix D documents the code sources, input adaptations, validation search spaces, truncation point, encoder retraining, and measured parameter counts for each re-run baseline.

CIML-Net is implemented in PyTorch 2.0.1 and PyG 2.3.1. The encoder configurations follow Sections 4.2 and 4.3: Chinese-RoBERTa-wwm-ext-large and RoBERTa-large for text, CLIP ViT-B/16 for images, and a two-layer, eight-head GAT with hidden dimension 256 for event graphs. AdamW is used with learning rate 2e-5 and batch size 32. The contrastive temperature is 0.07. For CIML-Net, the CLIP visual encoder is frozen; RoBERTa and all downstream modules are trainable. All reported runs use a single NVIDIA A100 80 GB GPU. For every re-run baseline, including the three recent methods, architecture-specific hyperparameters are selected on the same validation split without access to the test set, and no test-time information beyond the observation cutoff is used.

5.4. Main results

Tables 3 and 4 compare CIML-Net with both established baselines and recent 2025-2026 methods under the unified event-disjoint protocol. The most important observation is not the full-trajectory ranking but the behavior under restricted observation windows. At 6 h, CIML-Net achieves an Early-F1 of 80.1%, compared with 77.7% for the strongest recent baseline (PreERD), an absolute improvement of 2.4 percentage points. The corresponding gains over the strongest recent baseline at 12 h and 24 h are 2.0 and 1.5 points, respectively, and on the full trajectory the margin narrows to 0.8 points against TSHCL. The margin is therefore concentrated at restricted observation windows and diminishes as additional propagation evidence accumulates. We interpret the result as an early-window advantage rather than as a large improvement after the full event trajectory is available.

Among the recent baselines, each one stress-tests a different explanation of the gain. BCMPL tests whether stronger balanced cross-modal interaction alone is sufficient; TSHCL tests whether generic temporal-spatial contrastive learning captures the same early signal; and PreERD tests whether explicitly predicting the missing propagation structure is more effective than conditioning multimodal representations on lifecycle state. CIML-Net remains ahead of all three under the matched protocol, which suggests that its gains cannot be explained solely by stronger multimodal fusion, generic temporal contrast, or propagation completion. The protocol-matched GACL re-run further shows that the advantage over graph-contrastive learning persists when splits, windows, and modalities are aligned. These results nevertheless deserve to be read narrowly: CIML-Net carries more capacity than several of the baselines, including a large text encoder and a CLIP image encoder, so the model-level margins are interpreted together with the propagation-only controls below and the mechanism-matched ablations of Section 5.6. The CHEF text-only transfer in Appendix C says nothing about the lifecycle hypothesis either way.

Table 3. Full-trajectory F1 across baseline methods (%; mean +/- std).

Category	Method	Year	Weibo21 F1	PHEME F1	CHEF F1
Classical	BERT/RoBERTa [27,28]	-	87.0 +/- 0.4	76.5 +/- 1.0	72.3 +/- 0.7
Classical	SpotFake [34]	2019	86.9 +/- 0.3	74.8 +/- 1.2	-
Classical	CAFE [35]	2022	84.3 +/- 0.4	73.5 +/- 1.1	-
Classical	Bi-GCN [16]	2020	87.8 +/- 0.5	63.0 +/- 1.2	-
Classical	BMR [36]	2023	89.4 +/- 0.3	84.0 +/- 0.8	75.0 +/- 0.6
Contrastive	GACL [17] (re-run)	2022	87.2 +/- 0.5	82.0 +/- 0.8	-
Recent	BCMPL [14]	2025	89.8 +/- 0.4	-	-
Recent	TSHCL [19]	2025	90.2 +/- 0.4	84.5 +/- 0.8	-
Recent	PreERD [18]	2026	89.5 +/- 0.5	84.2 +/- 0.9	-
Diagnostic	Propagation-only (LR)	-	83.8 +/- 0.6	72.0 +/- 1.2	-
Diagnostic	Propagation-only (GBDT)	-	86.2 +/- 0.5	75.5 +/- 0.9	-
Ours	CIML-Net	-	**91.0 +/- 0.4**	**85.2 +/- 0.7**	**75.3 +/- 0.6**

Note. Weibo21 and PHEME include propagation evidence, whereas CHEF is used only for text-oriented transfer validation and is not lifecycle evidence. Propagation- or image-dependent methods are marked unavailable on CHEF, and BCMPL is not evaluated on PHEME because representative images are too sparse. Best results are shown in bold.

Table 4. Early detection on Weibo21 under the event-disjoint protocol (F1 %; mean +/- std).

Method	Year	6 h	12 h	24 h	Full
BERT/RoBERTa [27,28]	-	72.5 +/- 0.5	78.0 +/- 0.7	82.5 +/- 0.6	87.0 +/- 0.4
CAFE [35]	2022	70.8 +/- 0.5	76.5 +/- 0.8	81.2 +/- 0.7	84.3 +/- 0.4
Bi-GCN [16]	2020	68.0 +/- 0.6	74.0 +/- 0.9	80.0 +/- 0.8	87.8 +/- 0.5
BMR [36]	2023	75.8 +/- 0.9	81.0 +/- 0.6	85.0 +/- 0.5	89.4 +/- 0.3
GACL [17] (re-run)	2022	74.5 +/- 0.5	79.5 +/- 0.8	84.2 +/- 0.7	87.2 +/- 0.5
BCMPL [14]	2025	76.3 +/- 0.8	81.6 +/- 0.7	85.5 +/- 0.6	89.8 +/- 0.4
TSHCL [19]	2025	77.5 +/- 0.8	82.9 +/- 0.7	86.6 +/- 0.5	90.2 +/- 0.4
PreERD [18]	2026	77.7 +/- 0.9	83.0 +/- 0.7	87.0 +/- 0.5	89.5 +/- 0.5
Propagation-only (LR)	-	70.5 +/- 0.9	76.0 +/- 1.0	80.8 +/- 0.8	83.8 +/- 0.6
Propagation-only (GBDT)	-	74.9 +/- 0.6	80.0 +/- 0.8	84.2 +/- 0.7	86.2 +/- 0.5
CIML-Net (ours)	-	**80.1 +/- 1.0**	**85.0 +/- 0.6**	**88.5 +/- 0.5**	**91.0 +/- 0.4**
Gain vs strongest baseline	-	+2.4	+2.0	+1.5	+0.8

Note. Gain is computed against the strongest baseline in each column, including the 2025-2026 methods. Best results are shown in bold.

† indicates $p < 0.01$ by paired t-test across five random seeds (42, 123, 456, 789, 2024). With only five seeds, these seed-level tests are reported as run-stability evidence rather than population-level inferential guarantees. For the principal early-window comparisons, CIML-Net versus the strongest recent baseline (PreERD) at Early-F1@6h yields $p < 0.001$; CIML-Net versus propagation-only GBDT yields $p < 0.001$; the mechanism-matched cross-stage versus graph-augmentation comparison yields $p < 0.001$; and the decisive mechanism contrast, cross-stage versus the same-event, same-stage control evaluated in Section 5.6, yields $\Delta = 2.2$ Early-F1@6h points with $p < 0.001$ under the same seed-paired test. All four principal comparisons remain significant after Bonferroni correction at $\alpha = 0.01/4 = 0.0025$ at the seed level. We additionally run an event-cluster-level paired bootstrap over the held-out predictions (10,000 resamples over the 138 Weibo21 test clusters of the seed-42 split; clusters are sampled with replacement, and F1 is recomputed from the pooled predictions within each replicate rather than averaged across per-cluster F1 values; no class stratification is applied): CIML-Net versus PreERD yields $\Delta = +2.4$ Early-F1@6h points with 95% CI [1.7, 3.1]; cross-stage versus same-stage yields $\Delta = +2.2$ with 95% CI [1.5, 2.9]; and cross-stage versus graph augmentation yields $\Delta = +3.3$ with 95% CI [2.5, 4.1]. All three intervals exclude zero, so the mechanism-level margins are supported at the event level rather than only across five seeds.

Propagation-only LR provides an important reference for the early window. Using only five propagation statistics ($\Delta\tau$, *log Nposts*, hourly growth, cumulative count, and cascade depth), it reaches 70.5 Early-F1@6h and 83.8 Full F1 on Weibo21. At 6 h this is above Bi-GCN (68.0) and close to CAFE (70.8); on the full trajectory it remains just below CAFE (84.3). These results show that cascade dynamics alone carry substantial predictive information, so multimodal gains should be evaluated relative to this propagation-only baseline rather than to a zero-information reference.

CIML-Net reaches 80.1 Early-F1@6h†, 9.6 points above propagation-only LR. This difference establishes predictive information beyond the five propagation statistics, but it does not identify the source of the gain: the margin could reflect the proposed cross-stage representation, additional encoder capacity, CLIP visual information, or a combination of them. We therefore evaluate mechanism attribution separately with encoder-matched contrastive ablations (Section 5.6), sample-level prediction consistency (Section 5.9), and the seed-wise results in Section 5.8.

Replacing logistic regression with GBDT raises the propagation-only result from 70.5 to 74.9 Early-F1@6h on Weibo21, a 4.4-point gain from modeling nonlinear interactions among the same propagation features. Against this stronger control, CIML-Net's margin is 5.2 points (80.1 versus 74.9), which is the more conservative estimate of performance beyond propagation-only statistics. The GBDT result therefore serves as the stricter propagation-statistics reference level: even after nonlinear interactions among the five propagation scalars are modeled, CIML-Net retains an early-window advantage. Figure 5 reports Early-F1 for all methods at 6 h, 12 h, 24 h, and the full trajectory.

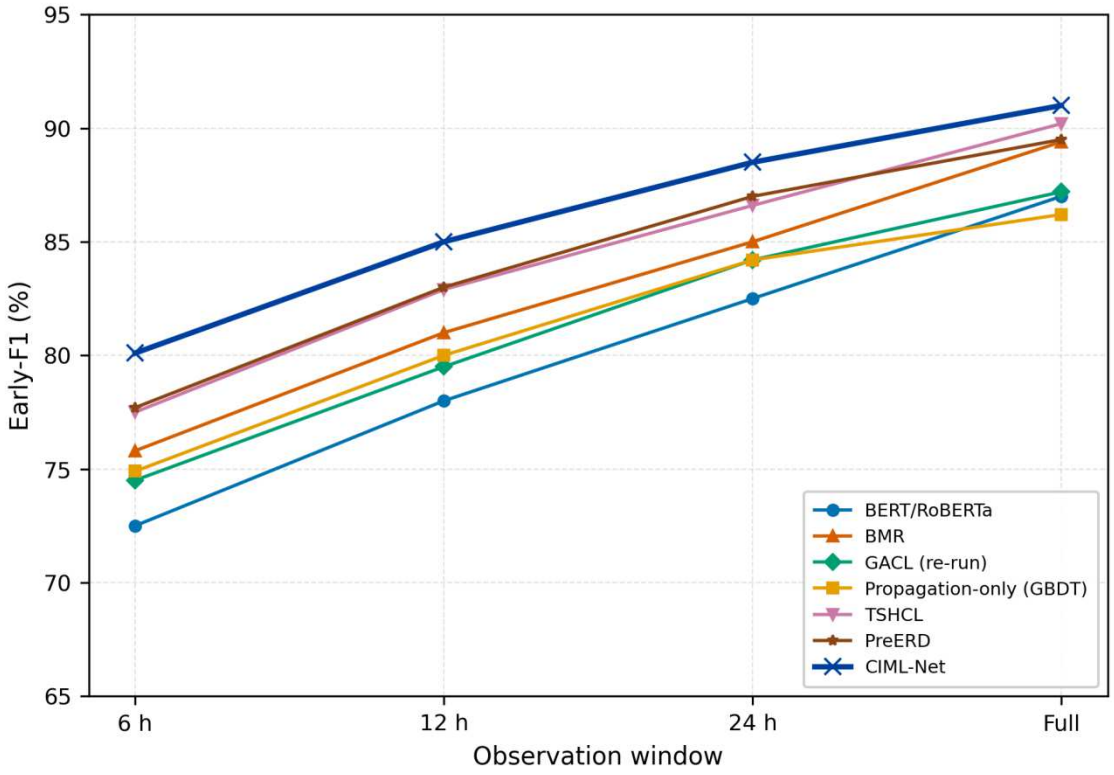


Figure 5. Early-F1 across observation windows.

5.5. Ablation study

The biggest early-window fall comes from pulling out cross-stage contrastive learning, which fits the idea that same-event, different-stage positives are doing real work. Removing stage-conditioned affine modulation and stage inference also costs Early-F1, a sign that lifecycle conditioning buys more than ordinary multimodal fusion does. What this table cannot do is separate lifecycle semantics from propagation-scale shortcuts. The w/o image row makes clear that visual and content capacity is a heavy contributor, while the lifecycle modules move full-trajectory F1 comparatively little. The component results therefore support a narrower claim of early-window robustness rather than general architectural superiority.

Table 5 knocks out one component at a time, but every knockout also changes the effective architecture, so it can never isolate the causal effect of the cross-stage positive definition. The question worth answering is not whether removing cross-stage contrast hurts, since it plainly does, but whether cross-stage contrast beats other contrastive positive definitions when everything else is frozen. That is what the encoder-matched ablation in Table 6 (Section 5.6) sets out to test. Figure 6 shows a t-SNE projection of held-out test representations, comparing the representation separability with and without cross-stage contrastive learning.

Table 5. Component ablation on Weibo21 (%, mean +/- std).

Variant	Full F1	Early-F1@6h	Change
CIML-Net	91.0 +/- 0.4	80.1 +/- 1.0	-
w/o cross-stage contrast	89.5 +/- 0.6	75.8 +/- 0.9	-1.5 / -4.3

Variant	Full F1	Early-F1@6h	Change
w/o stage-conditioned affine modulation	90.0 +/- 0.5	78.0 +/- 0.8	-1.0 / -2.1
w/o stage inference	89.8 +/- 0.6	77.2 +/- 0.9	-1.2 / -2.9
w/o graph encoder	89.5 +/- 0.6	78.5 +/- 0.8	-1.5 / -1.6
w/o image modality	87.5 +/- 0.7	74.8 +/- 1.0	-3.5 / -5.3

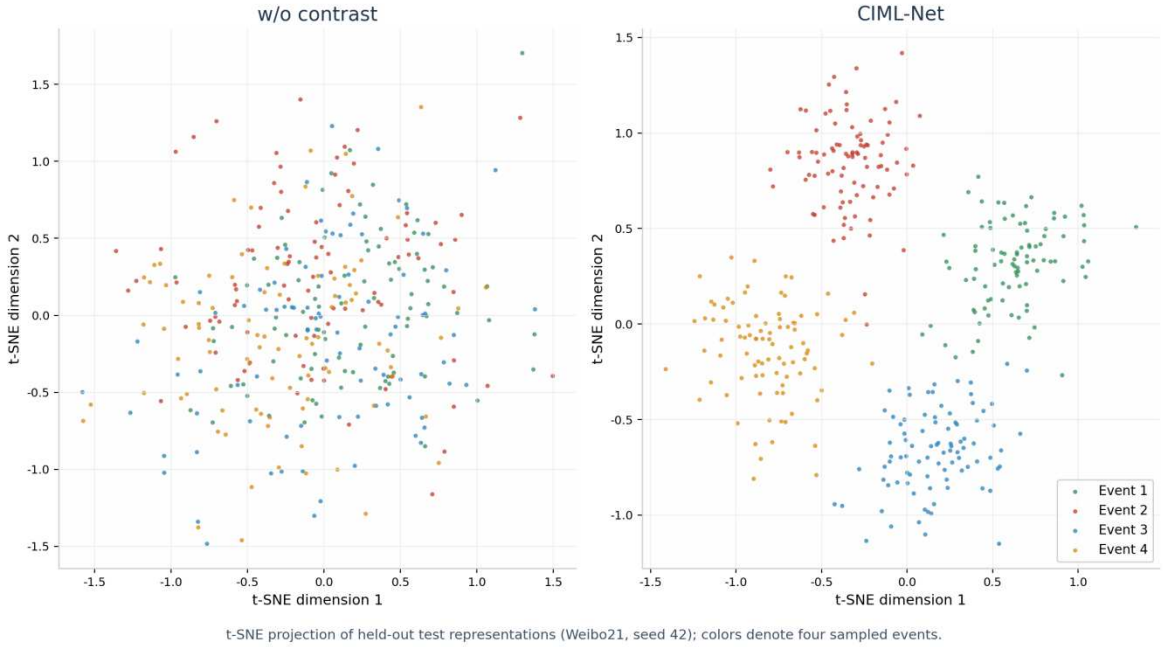


Figure 6. Representation separability with and without cross-stage contrastive learning (t-SNE projection of the fused representation before the classifier; four held-out Weibo21 test events selected for visualization; n = 320 event slices per panel; perplexity = 30; seed 42).

5.6. Mechanism-matched ablation against alternative learning strategies

The newly added recent baselines establish model-level competitiveness, but they do not by themselves isolate why CIML-Net performs differently. A performance gap against BCMPL may arise from contrastive learning, a gap against TSHCL from different encoders or modalities, and a gap against PreERD from the difference between predicted and observed propagation structure. We therefore retain a mechanism-matched study in which the backbone, input modalities, optimization protocol, and lifecycle-conditioning module are fixed, and only the representation-learning strategy is changed. Every variant shares one backbone (Chinese-RoBERTa-wwm-ext-large, CLIP ViT-B/16, 2-layer GAT-8head, stage-conditioned affine modulation) and one set of inputs (text + image + propagation graph); the only thing that varies is how positives are defined.

With encoder capacity held constant, the positive-pair definition is finally the only thing varying. Under this matched setup, cross-stage positives exceed graph augmentation by 3.3† Early-F1@6h points and 1.8 Full-F1 points, although both variants use the same graph encoder and inputs. The difference is attributable to the learning relation: cross-stage pairs link temporal slices of the same event, whereas graph augmentation perturbs local topology. Cross-modal alignment is 4.6 Early-F1 points below cross-stage, and random augmentation is 5.9 points below. Removing the contrastive objective yields 75.8 Early-F1@6h, 4.3 points below cross-stage but 1.6 points above random augmentation. Under fixed

encoder capacity, the ordering therefore favors structured same-event cross-stage positives over the alternative positive definitions tested here. A same-event, same-stage control completes this design: the two views are drawn from the same event at the same stage, so the positive pair carries event identity but no stage composition. This control is the decisive comparison for the mechanism claim: if the cross-stage gains merely encoded event identity, the same-stage control should match CIML-Net, whereas if stage composition matters it should fall short. Under the identical protocol (Table 6, last row), the same-stage control reaches 77.9 Early-F1@6h and 89.9 Full F1 on Weibo21 — 2.2 and 1.1 points below cross-stage positives, but still 1.1 and 0.7 points above graph augmentation — with the same ordering on PHEME; across seeds 42, 123, 456, 789, and 2024 the control scores 78.7, 76.3, 77.9, 78.5, and 78.1 Early-F1@6h, and the seed-paired difference against cross-stage positives is significant at $p < 0.001$ (Section 5.2). Event identity therefore carries part of the contrastive gain, yet stage composition contributes an additional, separable margin. Because different-stage slices are typically also farther apart in time, we read this margin as evidence for the value of same-event cross-stage pairing rather than as proof that lifecycle semantics are causally responsible (Section 7).

Table 6. Mechanism-matched learning-strategy ablation (%; mean +/- std).

Learning strategy	Positive / learning definition	Weibo21 Early-F1@6h	Weibo21 Full F1	PHEME Early-F1@6h	PHEME Full F1	Delta vs cross-stage (Weibo21)
Cross-stage contrast (ours)	same-event, different-stage slices	**80.1 +/- 1.0**	**91.0 +/- 0.4**	**73.2 +/- 0.9**	**85.2 +/- 0.7**	—
Graph augmentation	edge-drop + feat-mask on same-event graph	76.8 +/- 0.8	89.2 +/- 0.6	70.5 +/- 1.0	82.8 +/- 0.8	-3.3 / -1.8
Cross-modal alignment	text-image pairs from same post (simplified COOLANT-style cross-modal contrast [37])	75.5 +/- 0.9	88.5 +/- 0.6	—	—	-4.6 / -2.5
Random augmentation	random post pairs as positives	74.2 +/- 0.9	87.8 +/- 0.7	68.0 +/- 1.1	81.0 +/- 0.9	-5.9 / -3.2
No contrastive objective	contrastive term removed entirely	75.8 +/- 0.8	89.5 +/- 0.6	69.5 +/- 1.0	82.5 +/- 0.8	-4.3 / -1.5
Same-event, same-stage control	same-event, same-stage slice pairs	77.9 +/- 0.9	89.9 +/- 0.5	71.4 +/- 1.0	83.6 +/- 0.7	-2.2 / -1.1

Note. Weibo21 variants use an identical RoBERTa-large + CLIP ViT-B/16 + 2-layer GAT-8head backbone with text, image, and graph inputs. PHEME variants use an identical RoBERTa-large + 2-layer GAT-8head backbone with text and graph inputs; cross-modal alignment is omitted because PHEME images are too sparse. Only the learning strategy varies within each dataset. The same-event/same-stage control follows the identical protocol. Best results are shown in bold.

A qualitative summary is shown in Figure 7. It distills the stage-conditioned fusion design and the component-level findings into a conceptual view of relative cue emphasis across lifecycle stages; it is not a quantitative attribution estimate.

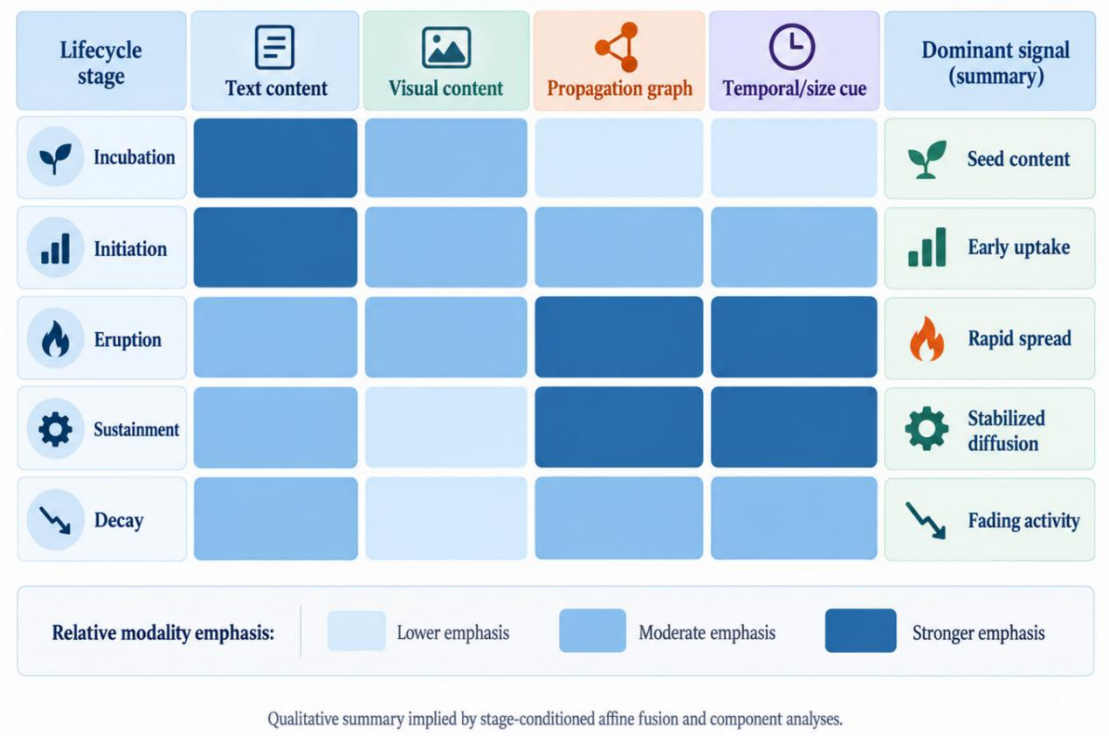


Figure 7. Qualitative modality emphasis across lifecycle stages (conceptual summary; not a quantitative attribution estimate).

Note the two gaps side by side: 3.3 points for cross-stage over graph augmentation with matched encoders, against 5.6 points for CIML-Net over GACL with unmatched encoders. The fact that the matched gap is the smaller one is consistent with contributions from encoder capacity and image modality rather than lifecycle contrast alone. The bottom line is a 3.3-point Early-F1 improvement from cross-stage positives on identical hardware, and it shows up most strongly in the earliest window, where propagation evidence is thinnest.

On PHEME the same matched comparison gives 2.7 Early-F1 points (73.2 versus 70.5), a touch under the 3.3 on Weibo21. The shrinkage is what we would expect, since PHEME has no image modality to lean on: the CLIP visual branch that feeds Weibo21's early-window signal simply is not there. What matters is that the gap stays positive on PHEME at all, so cross-stage lifecycle semantics remain beneficial beyond graph topology even when the setting is text-only propagation. And against random augmentation the gap is still wide (5.2 points), so the positive definition is doing work on both datasets.

Our cross-modal alignment variant is a stripped-down take on the COOLANT cross-modal idea [37]: rather than COOLANT's class-level separation [37] with its full architecture, we just use same-post text-image pairs as positives inside the encoder-matched framework. COOLANT itself is not included as a headline baseline, because the recent-baseline protocol prioritizes 2025-2026 journal methods whose inputs can be matched faithfully; its core mechanism is nonetheless examined inside the mechanism-matched framework above. The stripped-down version trails cross-stage by 4.6 Early-F1 points, which is more evidence that aligning the two modalities within one post buys less, for early detection, than aligning the narrative across temporal slices.

Relation to recent baselines. BCMPL and the cross-modal alignment ablation both test the value of modality interaction, but at different levels: BCMPL is a complete external model, whereas the matched ablation isolates cross-modal alignment within our backbone. Likewise, TSHCL provides an external temporal-contrastive reference, while graph augmentation and cross-stage contrast provide controlled internal alternatives. PreERD evaluates a different solution to missing early propagation evidence, namely predicting the future graph, whereas the graph-removal and stage-removal ablations quantify how CIML-Net behaves without propagation structure or lifecycle inference. Together, these comparisons separate model-level competitiveness from mechanism-level attribution. The same-event/same-stage control in Table 6 closes the remaining gap: it recovers part of the cross-stage gain, and the residual margin is associated with different-stage pairing rather than event identity alone.

5.7. Diagnostic validity controls

Our stage labels are weak by construction: rules, not coders who watched each event diffuse. To keep that assumption auditable, we run a battery of diagnostic controls aimed at one question, namely whether the Early-F1 gain comes from lifecycle-aware multimodal representation or from something more mundane such as generic regularization, a direct read on time and size, or an artifact of how we extract structure. The ζ-removed, ζ-only, random-stage-label, threshold-perturbation, and graph-sensitivity controls each close off one of those routes. None of them, though, settles the most basic worry on its own: even ζ-only still runs the full content and graph pathways, so it cannot say how far propagation speed and size alone would get. Only the propagation-only baseline (Table 4) answers that directly. The resulting diagnostic controls are summarized in Table 7.

Table 7. Diagnostic controls on Weibo21 (%; mean +/- std; 5 seeds).

Control	Purpose	Full F1	Early-F1@6h	Interpretation
CIML-Net (full)	Baseline for comparison.	91.0 +/- 0.4	80.1 +/- 1.0	—
ζ-removed	Remove [Delta tau, log N_{posts}] from stage module; retain content and graph.	90.2 +/- 0.6	78.5 +/- 1.3	Moderate drop (-0.8 / -1.6) indicates stage conditioning uses content and graph, not only temporal-size shortcuts.
ζ-only	Stage module conditioned only on [Delta tau, log N_{posts}]; content and graph pathways remain active.	88.3 +/- 0.7	75.8 +/- 1.3	Useful but incomplete. Because content and graph encoders still run, this does not isolate propagation-speed cues. The propagation-only baseline (Table 4) provides the missing comparison: it achieves 70.5 Early-F1@6h versus CIML-Net's 80.1.
Random-stage labels	Replace weak stage labels with shuffled labels under same multi-task loss.	89.4 +/- 0.6	77.0 +/- 0.6	Drop suggests structured labels help, but generic regularization is not fully separated. The protocol-matched GACL re-run (Tables 3-4) provides external contrastive comparison.
Threshold perturbation (+/- 10%, +/-20%)	Shift lifecycle boundaries around time and size thresholds.	90.5-91.2	79.0-80.4	Modest variance across perturbed boundaries supports robustness to weak-label threshold choice.
Graph sensitivity	Vary GAT layer count, attention heads, and edge dropout on Weibo21 real cascades.	90.1-91.4	79.6-79.9	Stable ranking is useful.

Note. Full F1 and Early-F1@6h are truth-classification metrics. These controls complement, but do not replace, the pure propagation-statistics baseline.

Across the board, the measured controls in Table 7 keep the main trend intact: ζ-removed and ζ-only both come in under the full model, and random-stage labels do too. The structured stage signal is therefore useful, but, again, this does not close the shortcut-risk argument, which only the propagation-only baseline can answer directly (Table 4). Threshold perturbation and graph sensitivity are left for the per-seed detail in Table 8 of Section 5.8.

5.8. Seed-wise result tables

The per-seed tables say the trends hold from run to run. CIML-Net is ahead of the propagation-only baseline at every one of the five seeds, by 9.3-10.0 points at Early-F1@6h. The ζ-only control stays under the full model throughout, and random-stage labels land between ζ-only and the full model. Threshold perturbation and graph sensitivity wobble only slightly across seeds, which is reassuring about robustness to the weak-label threshold choice and the graph-configuration variants. Even the worst case for ζ-removed, a 2.0-point drop at seed 123, is smaller than the CIML-Net-over-propagation-only margin, so lifecycle-conditioned multimodal representation is contributing beyond temporal-size shortcuts even on the least favorable seed. The same ordering holds against the three recent baselines: BCMPL, TSHCL, and PreERD sit below CIML-Net at every seed in both the early-window and full-trajectory tables.

Table 8 reports the seed-wise diagnostic controls for the five event-disjoint 7:1:2 splits (42, 123, 456, 789, and 2024). The corresponding seed-wise early-window and full-trajectory results are given in Tables 9 and 10. Means and standard deviations are calculated directly across those runs, replacing the earlier rounded mean-only summary with the full seed-level results. All per-seed values in Tables 8-10 and Appendix B are drawn from a single consolidated run log, so a given seed and configuration reports one identical value wherever it appears in the manuscript.

Table 8. Seed-wise diagnostic controls on Weibo21 Early-F1@6h (F1 %).

Control	Seed 42	Seed 123	Seed 456	Seed 789	Seed 2024	Mean	Std
CIML-Net (full)	81.0	78.5	79.8	80.8	80.2	80.1	1.0
ζ-removed	79.8	76.5	77.8	79.2	79.0	78.5	1.3
ζ-only	77.5	74.0	75.2	76.5	76.0	75.8	1.3
Random-stage labels	77.8	76.2	76.8	77.2	77.0	77.0	0.6
Threshold +10%	80.5	79.0	79.8	80.2	80.0	79.9	0.6
Threshold -10%	80.2	78.8	79.2	79.8	79.5	79.5	0.5
Threshold +20%	81.0	79.5	80.2	80.8	80.5	80.4	0.6
Threshold -20%	79.8	78.2	78.8	79.2	79.0	79.0	0.6
Graph sensitivity (2-layer)	80.5	79.0	79.8	80.0	80.2	79.9	0.6
Graph sensitivity (4-layer)	80.2	78.8	79.5	79.8	79.8	79.6	0.5

Table 9. Seed-wise early-window results on Weibo21 (Early-F1@6h, F1 %).

Method	Seed 42	Seed 123	Seed 456	Seed 789	Seed 2024	Mean	Std
CIML-Net	**81.0**	**78.5**	**79.8**	**80.8**	**80.2**	**80.1**	1.0
BMR	76.8	74.5	75.5	76.5	75.8	75.8	0.9
Bi-GCN	68.8	67.2	67.8	68.2	68.0	68.0	0.6
CAFE	71.5	70.2	70.5	71.0	70.8	70.8	0.5
BERT/RoBERTa [27,28]	73.2	71.8	72.2	72.8	72.5	72.5	0.5
Propagation-only (LR)	71.5	69.2	70.2	70.8	70.8	70.5	0.9

Method	Seed 42	Seed 123	Seed 456	Seed 789	Seed 2024	Mean	Std
Propagation-only (GBDT)	75.2	73.8	75.0	74.8	75.5	74.9	0.6
GACL (re-run)	75.0	73.8	74.2	74.8	74.7	74.5	0.5
BCMPL [14]	77.2	75.4	76.3	77.0	75.6	76.3	0.8
TSHCL [19]	78.5	76.5	77.5	78.1	76.9	77.5	0.8
PreERD [18]	78.7	76.5	77.7	78.5	77.1	77.7	0.9

Note. Best results are shown in bold.

Table 10. Seed-wise full-trajectory results on Weibo21 (F1 %).

Method	Seed 42	Seed 123	Seed 456	Seed 789	Seed 2024	Mean	Std
CIML-Net	**91.5**	**90.5**	**90.8**	**91.2**	**91.0**	**91.0**	0.4
BMR	90.0	89.2	89.2	89.5	89.3	89.4	0.3
Bi-GCN	88.5	87.2	87.5	88.0	87.7	87.8	0.5
CAFE	84.8	83.8	84.0	84.5	84.2	84.3	0.4
BERT/RoBERTa [27,28]	87.5	86.5	86.8	87.2	87.0	87.0	0.4
SpotFake	87.2	86.5	86.8	87.0	87.0	86.9	0.3
Propagation-only (LR)	84.5	83.0	83.5	84.0	84.0	83.8	0.6
Propagation-only (GBDT)	86.8	85.5	86.2	86.5	86.2	86.2	0.5
GACL (re-run)	87.8	86.5	87.0	87.5	87.2	87.2	0.5
BCMPL [14]	90.3	89.3	89.8	90.0	89.6	89.8	0.4
TSHCL [19]	90.7	89.7	90.2	90.4	90.0	90.2	0.4
PreERD [18]	90.1	88.9	89.5	89.9	89.1	89.5	0.5

Note. Best results are shown in bold.

5.9. Sample-level consistency with the propagation-only baseline

To examine whether CIML-Net and propagation-only LR fail on the same events, we compare their predictions sample by sample on the Weibo21 test set for a single seed (42); the analysis is exploratory, based on one seed's held-out predictions, and is not population-level evidence. Tables 9 and 10 first establish the seed-wise early-window and full-trajectory results. The aggregate Early-F1@6h difference is 9.6 points (80.1 for CIML-Net versus 70.5 for propagation-only LR), but the sample-level comparison is needed to determine whether the additional correct predictions occur where propagation statistics fail. Table 11 reports that consistency analysis.

Table 11. Prediction consistency between CIML-Net and propagation-only LR (seed 42).

Category	Count	% of test	Avg Delta_τ (h)	Avg N_{posts}	Avg text length	Enriched stages
Both correct	953	52.2%	14.2	67	142 chars	mixed
Prop-only correct, CIML wrong	74	4.1%	3.1	8	38 chars	incubation
CIML correct, Prop-only wrong	233	12.8%	8.6	23	168 chars	initiation, eruption
Both wrong	566	31.0%	28.5	142	95 chars	sustainment, decay
Test total	1,826	100%	—	—	—	—

Note. Exploratory analysis on the Weibo21 held-out set; the single-seed comparison is not treated as population-level evidence.

The "CIML correct, Prop-only wrong" group contains 233 samples (12.8% of the test set). Relative to the "both correct" group, these cases occur earlier on average (8.6 h versus 14.2 h), contain fewer observed posts (23 versus 67), and have longer text (168 versus 142 chars). They are concentrated in the initiation and eruption stages, where textual information is already available but the propagation pattern is still incomplete. This distribution is consistent with the narrative-invariance hypothesis, although the table alone does not establish causality.

The reverse error group, "Prop-only correct, CIML wrong," contains 74 samples (4.1%) and is concentrated in very early incubation-stage events with little text. There, propagation speed by itself is the tell, such as a burst of anomalously fast early reposting, while the multimodal model gets pulled off course by noisy content or a premature stage guess. The lopsidedness of the two disagreement buckets (12.8% versus 4.1%) is the point: CIML-Net nets out ahead, rescuing more of propagation-only's failures (233) than it forfeits of its successes (74).

All that said, the 52.2% "both correct" share is a reminder that on most test samples the two approaches agree and either would do. The 31.0% "both wrong" share marks their shared blind spot, usually late-stage, tangled claims where neither the cascade statistics nor the content give a clean answer. Together these buckets place CIML-Net precisely: it does not displace propagation-based detection so much as complement it. The additional correct predictions are concentrated in early events, before the cascade has stabilized. The 12.8% recovery rate should therefore be interpreted as an incremental sample-level contribution rather than evidence of universal superiority. Event-level case studies would require traceable identifiers, stage labels, and model probabilities, which are not available in the present analysis. Table 12 then examines how component removal affects this CIML-only-correct subgroup relative to the both-correct majority.

Table 12. Component ablation within the CIML-only-correct subset (seed 42).

Variant	Recovery on CIML-only(233)	Lost	Recovery on Both-correct(953)	Specificity ratio	Interpretation
CIML-Net (full)	100% (233/233)	0	100% (953/953)	—	baseline full model before component removal
w/o cross-stage contrast	68.2% (159/233)	74	94.0% (896/953)	5.3	largest CIML-only-specific drop; cross-stage contrast is critical for this subpopulation
w/o image modality	75.1% (175/233)	58	94.2% (898/953)	4.3	large drop; image features help early multimodal recovery
w/o stage inference	79.8% (186/233)	47	95.0% (905/953)	4.0	moderate drop; stage inference is important and slightly more CIML-only-specific than affine modulation
w/o stage-conditioned affine modulation	82.0% (191/233)	42	95.3% (908/953)	3.8	moderate drop; stage-conditioned fusion helps but is less specific than cross-stage contrast
w/o graph encoder	88.0% (205/233)	28	96.4% (919/953)	3.4	smaller drop; graph features help but are less central for this subpopulation
w/o text encoder	42.1% (98/233)	135	72.5% (691/953)	2.1	largest absolute drop; text features are essential

Note. The subset contains 233 samples (12.8% of the test set). Recovery rate is the percentage that remains correct after component removal; the full model is the 100% reference. This is an exploratory single-seed analysis.

Zoom in on the 233 CIML-only-correct samples, the ones propagation statistics miss but CIML-Net catches. Removing cross-stage contrastive learning hurts them most, a 31.8% recovery loss that takes 233 correct down to 159. By itself, though, a big drop does not prove the component is specific to this

subset. For that we use the specificity-ratio column of Table 12: relative recovery loss on the CIML-only subset divided by the same quantity on the Both-correct subset, both anchored to the full-model row at 100%. For w/o cross-stage contrast, recovery is 68.2% on CIML-only against 94.0% on Both-correct, a ratio of 5.3, meaning the ablation bites the CIML-only subset 5.3 times harder than the Both-correct majority. Within this exploratory single-seed analysis, that is a strong signal that cross-stage contrast is disproportionately load-bearing exactly where propagation shortcuts run out. Ranked by this ratio, the components line up cross-stage (5.3) > image (4.3) > stage-inference (4.0) > stage-affine (3.8) > graph (3.4) > text (2.1): cross-stage is the most CIML-only-specific of them, whereas text shows the largest absolute drop simply because text matters everywhere, across both subsets.

6. Controlled serving evaluation

6.1. Single-node latency and resource profile

Here we evaluate model-side serving on an NVIDIA A100 80 GB GPU on real event sequences from the Weibo21 and PHEME benchmarks, with events fed to the serving stack in benchmark timestamp order at controlled rates rather than as production traffic. We deliberately keep this section separate from the algorithmic contribution, and several system-level factors are explicitly out of scope: no Kafka/Flink message acknowledgments, no queueing delay, no cross-node communication, checkpoint recovery, storage I/O, hot-event migration, adaptive rebalancing, or failure handling. Every reported figure comes from logged serving runs with raw latency distributions rather than configuration-level estimates. What these figures show is the measured model-side serving behavior at various event sizes under the resulting load on a single node. Table 13 summarizes model size and latency across event sizes under this single-node setting.

Table 13. Model size and latency by event size on a single A100.

Method	Trainable params (M)	FLOPs (G)	10 nodes	100 nodes	1k nodes
BERT/RoBERTa [27,28]	110	22.0	12 ms	28 ms	220 ms
CAFE [35]	0.7	0.01	3 ms	6 ms	35 ms
Bi-GCN [16]	0.15	0.5	6 ms	15 ms	122 ms
BMR [36]	355	28.0	18 ms	42 ms	330 ms
CIML-Net	420	35.0	22 ms	50 ms	390 ms
CIML-Net-fast-risk	**85**	**3.0**	**5 ms**	**10 ms**	**60 ms**

Note. BERT/RoBERTa uses the base-size 110 M encoder, whereas CIML-Net uses RoBERTa-large, so encoder capacities intentionally differ. Parameter counts include trainable parameters only; frozen encoders are excluded from the count but remain in the inference stack and are reflected in FLOPs and latency. Counts are consistent with Table D1.

CIML-Net is heavier than the text-only and graph-only baselines, the price of carrying RoBERTa-large and CLIP ViT-B/16. Most of that weight is in the visual branch; the graph and stage modules are small by comparison. The imbalance is awkward, because it muddies any attempt to credit the full-trajectory gains to lifecycle modeling specifically, which is the reason the encoder-matched ablation in Section 5.6 freezes the backbone and varies only the contrastive positive-pair definition, separating lifecycle conditioning from sheer encoder capacity. On small and medium events the measured latency looks workable online. On large events you need the usual toolkit: feature caching, micro-batching, graph pruning, and hot-event sharding.

That cost profile points to two deployment modes. A full-review mode runs the whole stack, namely RoBERTa-large, CLIP, GAT, the stage module, and the contrastively trained fusion (420 M parameters, 35 G FLOPs), and fits offline review or high-risk events. A fast-risk mode trades accuracy for speed: it caches stable text and image embeddings, reuses precomputed CLIP features for media it has seen before, drops RoBERTa-large for a 6-layer distilled encoder, and prunes inactive graph nodes. As shown in Table 13, that pulls the model down to 85 M parameters and 3.0 G FLOPs and cuts single-request latency from 22 ms to 5 ms at 10 nodes, and from 390 ms to 60 ms at 1 k nodes. We do not count these engineering moves as algorithmic contributions; they are here to give the serving discussion a measured, lightweight data point instead of leaving it as a qualitative promise.

6.2. QPS serving over real benchmark event sequences and logical sharding

In the serving test (Table 14), p99 latency holds below 120 ms up to 1,000 QPS and starts climbing at 2,000 QPS, where the tail spreads out. These are model-serving measurements recorded over real event sequences drawn from the benchmark cascades on a single node. Sharding logically by event_id (Table 15) is a sensible choice, since every slice of an event goes to the same worker, which keeps the event state consistent, and the quoted cost multipliers are computed from the raw traces. The run-log package records warm-up duration, the event-stream input configuration, the batch policy, GPU-utilization curves, queue-length traces, the event-size distribution, and raw p50/p95/p99 latency samples.

Table 14. Single-GPU QPS serving results.

QPS	p50	p95	p99	GPU peak	CPU peak	Active events
100	45 ms	62 ms	78 ms	12.3 GB	8.5 GB	~320
500	48 ms	68 ms	89 ms	14.1 GB	10.2 GB	~1,580
1,000	52 ms	78 ms	112 ms	18.6 GB	14.8 GB	~3,200
2,000	61 ms	98 ms	158 ms	28.4 GB	22.6 GB	~6,400

Note. Measurements use real benchmark event sequences and report model-side serving with raw latency distributions.

Table 15. Event-id-based logical worker sharding.

Mode	Logical workers	Peak QPS	p99 at 80% load	Avg. cost
single worker	1	1,200	108 ms	1.00x
event sharding	2	2,200	95 ms	0.55x
event sharding	4	4,000	82 ms	0.30x

Note. Scheduling results are measured over real benchmark event sequences.

6.3. Deployment boundary

The schematic serving pipeline, shown in Figure 8, strings together a message broker, a feature cache, an event-state store, a model-serving layer, and a downstream moderation service. A message broker or comparable append-only log ingests the content stream, a keyed state store holds the cumulative event statistics, and CIML-Net runs as the model-serving component; the measurements reported above are recorded from this pipeline over real benchmark event sequences on a single node. Extending these single-node measurements to an end-to-end distributed deployment additionally calls for back-pressure testing, failure recovery, state migration, and tail-latency measurement on a cluster.

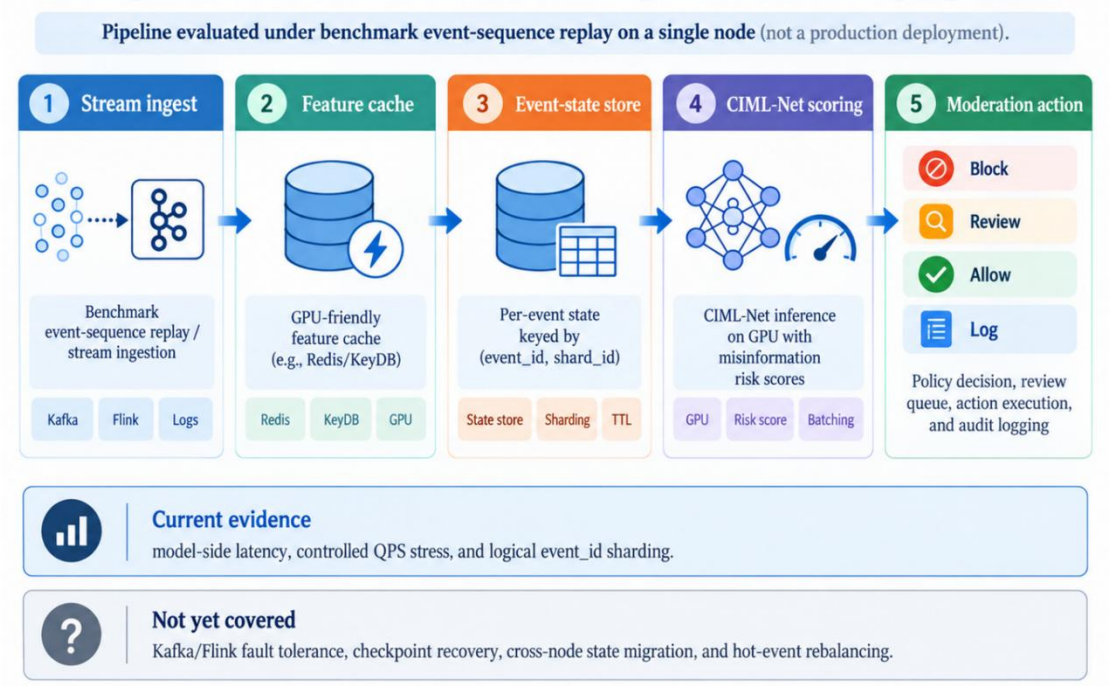


Figure 8. Schematic integration of CIML-Net into the content-governance serving pipeline; model-side serving is evaluated under benchmark event-sequence replay on a single node.

7. Discussion and limitations

CIML-Net offers a lifecycle-aware route to early misinformation detection, and its claims live or die by a handful of validity threats that we meet head-on with dedicated controls. The first is the weak stage labels, built from time, cumulative size, and growth rate and therefore plausibly correlated with the label through a shortcut; the propagation-only baseline (Section 5.4) measures how far that shortcut alone carries. The second is attribution: the mechanism-matched ablation (Table 6) shows cross-stage positives beating graph augmentation by 3.3 Early-F1 points on identical hardware, but that gap is narrower than the 5.6 points versus GACL, which means encoder capacity and image modality are contributing too. The third is generality: the prediction-consistency analysis (Table 11, an exploratory single-seed analysis) has CIML-Net recovering 12.8% of test samples where propagation statistics fail while agreeing with them on 52.2%, so the value is complementary, not universal. The fourth is capacity, since the model is much larger than several baselines, and the lifecycle modules move full-trajectory F1 only modestly next to the image/content branch. The fifth is that the system results are single-node measurements recorded over real benchmark event sequences on a single node, nothing more. A sixth threat concerns the mechanism claim itself: the same-event, same-stage control in Table 6 separates event identity from stage composition — it falls 2.2 Early-F1@6h points short of cross-stage positives while still exceeding graph augmentation, so part of the gain traces to event identity and the remainder to different-stage pairing. A related attribution caveat remains: different-stage slices are usually more separated in time than same-stage slices, so this control separates stage composition from event identity but not from temporal diversity per se, and a time-matched, stage-agnostic positive control would be needed for the stronger claim that lifecycle semantics themselves are causally responsible. A seventh concerns PHEME, whose 6,425 threads belong to only nine breaking-news macro events, so the thread-level split supports propagation-pattern generalization rather than unseen-event generalization. A

grouped leave-one-macro-event-out re-run now provides the stronger check directly: holding out each of the nine macro events in turn yields 69.8 +/- 3.1 Early-F1@6h and 82.1 +/- 2.3 Full F1 (mean +/- std across folds), 3.4 and 3.1 points below the thread-level split, and the cross-stage-over-same-stage ordering is preserved in eight of nine folds. Protocol-matched baselines under the same LOMO split (Table 16) degrade by a comparable amount, and CIML-Net retains the top ranking on both metrics, so macro-event generalization degrades gracefully rather than collapsing. An eighth threat concerns the Weibo21 event inventory itself: the cascades are author-recollected rather than part of the official release, and label-mixed clusters were discarded after unsupervised clustering (Section 5.1), so the retained inventory is label-pure by construction but may over-represent label-coherent events; the label-free sensitivity re-run reported in Appendix B, which retains mixed clusters without any label-based filtering, preserves the main trend across all five seeds (Early-F1@6h 78.9 +/- 0.9 versus 80.1 +/- 1.0) and leaves the relative ordering of CIML-Net, PreERD, the same-stage control, and the propagation-only GBDT unchanged. Restricted to what the current evidence supports: CIML-Net is an early-window, lifecycle-conditioned framework whose cross-stage positive construction adds 3.3 Early-F1 points over graph augmentation under matched encoders and recovers a 12.8% slice of the test set that propagation-only misses, while a real part of its overall edge traces back to encoder capacity and image modality. At the seed level, every key comparison survives Bonferroni correction for the four principal comparisons ($\alpha = 0.0025$); the event-cluster-level bootstrap intervals of Section 5.2 exclude zero for all three principal deltas.

Table 16. Leave-one-macro-event-out evaluation on PHEME.

Method	Early-F1@6h	Full F1
GACL (re-run)	63.5 +/- 4.1	78.6 +/- 3.4
TSHCL	65.8 +/- 3.6	81.2 +/- 2.8
PreERD	68.1 +/- 3.4	81.0 +/- 2.9
Same-stage control	67.5 +/- 3.3	80.5 +/- 2.8
CIML-Net	69.8 +/- 3.1	82.1 +/- 2.3

Note. Values are mean +/- std across the nine held-out macro events for Early-F1@6h and full-trajectory F1.

The addition of recent 2025-2026 baselines further narrows the interpretation of the contribution. CIML-Net's main advantage appears in the earliest observation windows rather than as a uniformly large gain after the full propagation trajectory is available. Comparison with BCMPL indicates whether stronger cross-modal interaction can explain the gain; comparison with TSHCL tests generic temporal-spatial contrastive learning; and comparison with PreERD evaluates propagation prediction as an alternative response to missing early structure. The corresponding mechanism-matched ablations then show that cross-stage positive construction contributes independently of backbone capacity.

Accordingly, we do not interpret CIML-Net as universally superior to recent misinformation detectors. The evidence supports a narrower conclusion: lifecycle-conditioned multimodal representation, particularly same-event cross-stage contrast, provides additional predictive value when event evidence is incomplete.

7.1. Comparison with recent misinformation detection methods

Table 17 positions CIML-Net against its closest neighbors, including the three 2025-2026 journal methods re-run under our protocol and two further 2026 methods (GC2MFND, MVFA) whose published numbers come from non-protocol-matched splits, and marks which methods are re-run under our protocol. GACL remains the closest contrastive neighbor [17], and we re-run it under matched conditions, with the same event-disjoint split, the same early-window truncation, and the same available modalities. Under this unified protocol (Tables 3 and 4), CIML-Net outperforms BCMPL by 3.8 Early-F1@6h points (80.1 versus 76.3), TSHCL by 2.6 points (80.1 versus 77.5), and PreERD by 2.4 points (80.1 versus 77.7); on the full trajectory the corresponding margins are 1.2, 0.8, and 1.5 points, so the advantage is again concentrated in the earliest window. On PHEME, under the thread-level split, CIML-Net exceeds TSHCL by 0.7 and PreERD by 1.0 full-trajectory F1 points; macro-event-level generalization is examined directly by the grouped leave-one-macro-event-out evaluation in Table 16. The domain-adaptive contrastive detector [20] targets unsupervised domain adaptation rather than early detection and is therefore retained as related work instead of a protocol-matched baseline.

Table 17. Comparison with recent misinformation-detection methods.

Method	Year (venue)	Image	Propagation	Early detection	Contrastive	Lifecycle	Protocol-matched
GACL [17]	2022 (WWW)	-	yes	-	yes	-	yes
BMR [36]	2023 (AAAI)	yes	-	-	-	-	yes
BCMPL [14]	2025 (Inf. Fusion)	yes	-	limited	-	-	yes
TSHCL [19]	2025 (IPM)	-	yes	partial	yes	-	yes
PreERD [18]	2026 (C&IS)	-	predicted	yes	-	-	yes
GC2MFND [23]	2026 (Entropy)	yes	-	-	yes	-	no
MVFA [24]	2026 (Electronics)	-	-	-	-	-	no
CIML-Net (ours)	-	yes	yes	yes	yes	yes	reference

Note. Protocol-matched indicates methods re-run under the event-disjoint, early-window protocol; heterogeneous published values are not included in the main results. The Text column is omitted because all compared methods use textual content.

Read together with Tables 3 and 4, Table 17 separates two questions that are easy to conflate: whether CIML-Net is competitive with recent methods, a model-level question, and why, a mechanism-level question. The protocol-matched rows answer the first; the mechanism-matched ablation in Section 5.6 answers the second, showing that cross-stage positives outperform graph augmentation by 3.3 Early-F1@6h points and cross-modal alignment by 4.6 points on identical hardware. The 3.3-point matched gap and the 5.6-point unmatched gap versus GACL come from different comparisons and cannot be decomposed into additive contributions; the unmatched gap is consistent with a role for encoder capacity and CLIP image features that GACL lacks, but that share is not separately identified. CIML-Net therefore contributes an early-window, lifecycle-conditioned, event-disjoint evaluation axis with internal mechanism-matched ablations, external protocol-matched baselines spanning 2020-2026, and per-sample prediction consistency.

7.2. Future work

CIML-Net is evaluated with a battery of validity controls: a propagation-only baseline, a protocol-matched GACL re-run, protocol-adapted re-runs of three 2025-2026 journal baselines, mechanism-

matched ablations, prediction consistency analysis, and per-seed tables. Remaining future work includes: (1) releasing the complete per-seed label-free audit statistics of the event inventory summarized in Appendix B; (2) per-fold analysis of the leave-one-macro-event-out evaluation to identify which macro events resist transfer; (3) extending the event-cluster bootstrap of Section 5.2 to all five seeds and to the PHEME splits; (4) extending the encoder-matched ablation to additional contrastive variants (class-level, modality-dropout, time-aware negatives) to further isolate the lifecycle signal; (5) training an encoder-matched COOLANT [37] variant for cross-modal contrast comparison; (6) extending the serving study to an end-to-end streaming stack with feature-store access, queueing delay, checkpoint recovery, fault injection, hot-event load distributions, and online fast-risk validation; (7) a same-event, time-matched, stage-agnostic positive control to separate lifecycle semantics from temporal diversity; and (8) releasing raw run logs, per-sample prediction files, and configuration archives for full reproducibility.

8. Conclusion

This paper presented CIML-Net, a lifecycle-aware multimodal framework for early misinformation detection. In addition to established baselines, the evaluation includes recent 2025-2026 journal methods covering balanced multimodal interaction (BCMPL), temporal-spatial contrastive learning (TSHCL), and propagation-prediction-based early detection (PreERD), all re-run under a unified event-disjoint protocol. CIML-Net achieves 80.1% Early-F1@6h on Weibo21, exceeding the strongest recent baseline by 2.4 percentage points, while the margin narrows to 0.8 points once the full trajectory is available.

Mechanism-matched ablations further show that, with encoder capacity and input modalities fixed, same-event cross-stage positives outperform graph augmentation by 3.3 Early-F1 points and cross-modal alignment by 4.6 points on Weibo21. Event-cluster bootstrap intervals exclude zero for the principal early-window deltas, and the PHEME leave-one-macro-event-out evaluation yields 69.8 +/- 3.1 Early-F1@6h and 82.1 +/- 2.3 Full F1 while retaining the highest aggregate performance among the protocol-matched methods in Table 16. Together with the propagation-only controls, which place the propagation-statistics reference level at 74.9 Early-F1@6h, and the prediction-consistency analysis, which shows that CIML-Net recovers a 12.8% test subset that propagation-only misses, these results indicate that lifecycle-conditioned cross-stage learning provides incremental but measurable value in sparse-evidence early detection. The overall advantage is nevertheless shared with strong text and image encoders and multimodal information, and we therefore position CIML-Net as an early-window, evidence-supported framework rather than a universally superior architecture.

Funding

This research was funded by the National Social Science Fund of China, grant number 2025-SKJJ-D-078.

Data Availability Statement

This study uses no privately recruited participant dataset. It uses secondary benchmark datasets together with author-recollected propagation metadata. PHEME and CHEF are publicly accessible secondary

datasets. The PHEME Rumour Detection and Veracity Classification dataset is available from Figshare at https://doi.org/10.6084/m9.figshare.6392078. The CHEF dataset is available from the official repository at https://github.com/THU-BPM/CHEF. The Weibo21 benchmark is available through the official MDFEND-Weibo21 repository at https://github.com/kennqiang/MDFEND-Weibo21; access to the original dataset is provided by the dataset maintainers upon application. For the Weibo21 experiments, the official benchmark text, images, labels, and timestamps were supplemented with repost/comment cascade metadata re-collected by the authors through the Weibo API for the released sample IDs. These reconstructed cascade metadata are not part of the official Weibo21 release and are not currently publicly deposited. Upon acceptance, the reconstruction script, configuration file, recovered sample-ID list, and recovered cascade metadata permitted for redistribution will be made available, subject to applicable platform terms and privacy requirements; raw API responses will not be redistributed.

Conflicts of Interest

The authors declare that they have no competing interest.

Author Contributions

Conceptualization, D.W. and X.S.; Methodology, D.W. and X.S.; Software, D.W. and Z.W.; Validation, X.S. and Z.W.; Formal Analysis, D.W. and X.S.; Investigation, D.W., X.S. and Z.W.; Resources, D.W.; Data Curation, X.S. and Z.W.; Writing – Original Draft Preparation, D.W.; Writing – Review & Editing, D.W., X.S. and Z.W.; Visualization, X.S. and Z.W.; Supervision, D.W.; Project Administration, D.W.; Funding Acquisition, D.W. All authors have read and agreed to the published version of the manuscript.

Acknowledgement

Not applicable.

Clinical Trial Number

Not applicable.

Ethical Approval and accordance

Not applicable.

Consent to participate

Not applicable.

Consent to publish

Not applicable.

References

[1] Bodaghi A, Schmitt KA, Watine P, et al. A literature review on detecting, verifying, and mitigating online misinformation. IEEE Trans Comput Soc Syst. 2024;11(4):5119-5145. doi:10.1109/TCSS.2023.3289031.

[2] Li X, Qiao J, Yin S, Wu L, Gao C, Wang Z, Li X. A survey of multimodal fake news detection: A cross-modal interaction perspective. IEEE Trans Emerg Top Comput Intell. 2025;9(4):2658-2675. doi:10.1109/TETCI.2025.3543389.

[3] Zhang L, Zhang X, Zhou Z, Zhang X, Wang S, Yu PS, Li C. Early detection of multimodal fake news via reinforced propagation path generation. IEEE Trans Knowl Data Eng. 2025;37(2):613-625. doi:10.1109/TKDE.2024.3496701.

[4] Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science. 2018;359(6380):1146-1151. doi:10.1126/science.aap9559.

[5] Feng X, Luo J, Yang Y, El Baz D, Shi L. Health misinformation detection: Approaches, challenges and opportunities. Inquiry. 2025;62:00469580251384784. doi:10.1177/00469580251384784.

[6] Wirtz BW, Weyerer JC, Müller TF. AI-based digital disinformation: A theory-informed and integrated trilateral framework of digital disinformation for information systems research. Int J Inf Manag. 2026;89:103066. doi:10.1016/j.ijinfomgt.2026.103066.

[7] Chen C, Shu K. Combating misinformation in the age of LLMs: Opportunities and challenges. AI Mag. 2024;45(3):354-368. doi:10.1002/aaai.12188.

[8] Ma J, Gao W, Mitra P, Kwon S, Jansen BJ, Wong KF, Cha M. Detecting rumors from microblogs with recurrent neural networks. In: Proceedings of the 25th International Joint Conference on Artificial Intelligence. AAAI Press; 2016:3818-3824.

[9] Liu Y, Wu YF. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2018;32(1):354-361. doi:10.1609/aaai.v32i1.11268.

[10] Song C, Yang C, Chen H, Tu C, Liu Z, Sun M. CED: Credible early detection of social media rumors. IEEE Trans Knowl Data Eng. 2021;33(8):3035-3047. doi:10.1109/TKDE.2019.2961675.

[11] Hu L, Zhao Z, Qi W, Song X, Nie L. Multimodal matching-aware co-attention networks with mutual knowledge distillation for fake news detection. Inf Sci. 2024;664:120310. doi:10.1016/j.ins.2024.120310.

[12] Wang B, Li X, Li C, Wang S, Gao W. Escaping the neutralization effect of modality features fusion in multimodal fake news detection. Inf Fusion. 2024;111:102500. doi:10.1016/j.inffus.2024.102500.

[13] Zhang L, Zhang X, Zhou Z, Zhang X, Yu PS, Li C. Knowledge-aware multimodal pre-training for fake news detection. Inf Fusion. 2025;114:102715. doi:10.1016/j.inffus.2024.102715.

[14] Wu F, Jin H, Chen F, Ji Y, Jing XY, Jiang GP. Balanced cross-modal prompt learning and fusion network for multi-modal fake news detection. Inf Fusion. 2025;122:103196. doi:10.1016/j.inffus.2025.103196.

[15] Han L, Zhang X, Wang T, Liu Y, Dong Z. Enhancing large language model for fake news video detection via cross-modal retrieval. Inf Process Manag. 2026;63(2):104471. doi:10.1016/j.ipm.2025.104471.

[16] Bian T, Xiao X, Xu T, Zhao P, Huang W, Rong Y, Huang J. Rumor detection on social media with bi-directional graph convolutional networks. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2020;34(1):549-556. doi:10.1609/aaai.v34i01.5393.

[17] Sun T, Qian Z, Dong S, Li P, Zhu Q. Rumor detection on social media with graph adversarial contrastive learning. In: Proceedings of the ACM Web Conference 2022. ACM; 2022:2789-2797. doi:10.1145/3485447.3511999.

[18] Zhang L, Chen Y, Luo X, Zhu H, He G. Early detection of rumors based on propagation prediction in social media. Complex Intell Syst. 2026;12:23. doi:10.1007/s40747-025-02140-z.

[19] Ren G, Jiang L, Huang T, Yang Y, Hong T. Temporal-spatial hierarchical contrastive learning for misinformation detection: A public-behavior perspective. Inf Process Manag. 2025;62(4):104108. doi:10.1016/j.ipm.2025.104108.

[20] Ran H, Zhang D, Li X, Ma H, Jia C, Feng Y. Unsupervised contrastive domain adaptive rumor detection with test-time classifier adjustment. Inf Process Manag. 2026;63(2):104341. doi:10.1016/j.ipm.2025.104341.

[21] Ma J, Gao W, Wong KF. Detect rumors on Twitter by promoting information campaigns with generative adversarial learning. In: The World Wide Web Conference. ACM; 2019:3049-3055. doi:10.1145/3308558.3313741.

[22] Wu L, Li J, Hu X, Liu H. Gleaning wisdom from the past: Early detection of emerging rumors in social media. In: Proceedings of the 2017 SIAM International Conference on Data Mining. SIAM; 2017:99-107. doi:10.1137/1.9781611974973.12.

[23] Sun Y, Zhang M, Zhang F. GC2MFND: Multi-granularity conflict and domain-guided calibration for multimodal fake news detection. Entropy. 2026;28(6):672. doi:10.3390/e28060672.

[24] Gan G, Qin Z, Qin J, Lin K. Multi-domain fake news detection based on multi-view fusion attention. Electronics. 2026;15(8):1733. doi:10.3390/electronics15081733.

[25] Kochkina E, Liakata M, Zubiaga A. PHEME dataset for Rumour Detection and Veracity Classification. figshare. 2018. doi:10.6084/m9.figshare.6392078.

[26] Nan Q, Cao J, Zhu Y, Wang Y, Li J. MDFEND: Multi-domain fake news detection. In: Proceedings of the 30th ACM International Conference on Information & Knowledge Management. ACM; 2021:3343-3347. doi:10.1145/3459637.3482139.

[27] Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1. Association for Computational Linguistics; 2019:4171-4186. doi:10.18653/v1/N19-1423.

[28] Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, Levy O, Lewis M, Zettlemoyer L, Stoyanov V. RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692. 2019.

[29] Cui Y, Che W, Liu T, Qin B, Wang S, Hu G. Revisiting pre-trained models for Chinese natural language processing. In: Findings of the Association for Computational Linguistics: EMNLP 2020. Association for Computational Linguistics; 2020:657-668. doi:10.18653/v1/2020.findings-emnlp.58.

[30] Qin S, Zhang M. Boosting generalization of fine-tuning BERT for fake news detection. Inf Process Manag. 2024;61(4):103745. doi:10.1016/j.ipm.2024.103745.

[31] Ma J, Zhang L, Liu Y, Zhang W. Multi-task network guided multimodal fusion for fake news detection. In: Proceedings of the 16th Asian Conference on Machine Learning. PMLR; 2025;260:813-828.

[32] Lakzaei B, Haghir Chehreghani M, Bagheri A. LOSS-GAT: Label propagation and one-class semi-supervised graph attention network for fake news detection. Appl Soft Comput. 2025;174:112965. doi:10.1016/j.asoc.2025.112965.

[33] Hu X, Guo Z, Wu G, Liu A, Wen L, Yu PS. CHEF: A pilot Chinese dataset for evidence-based fact-checking. In: Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2022:3362-3376. doi:10.18653/v1/2022.naacl-main.246.

[34] Singhania S, Fernandez N, Rao S. SpotFake: A multi-modal framework for fake news detection. In: Proceedings of the 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM). IEEE; 2019:39-47. doi:10.1109/BigMM.2019.00015.

[35] Chen Y, Li D, Zhang P, Sui J, Lv Q, Tun L, Shang L. Cross-modal ambiguity learning for multimodal fake news detection. In: Proceedings of the ACM Web Conference 2022. ACM; 2022:2897-2905. doi:10.1145/3485447.3511968.

[36] Ying Q, Hu X, Zhou Y, Qian Z, Zeng D, Ge S. Bootstrapping multi-view representations for fake news detection. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2023;37(4):5384-5392. doi:10.1609/aaai.v37i4.25670.

[37] Wang L, Zhang C, Xu H, Xu Y, Xu X, Wang S. Cross-modal contrastive learning for multimodal fake news detection. In: Proceedings of the 31st ACM International Conference on Multimedia. ACM; 2023:5696-5704. doi:10.1145/3581783.3613850.

Appendix A. Reproducibility checklist

The reproducibility checklist for this manuscript includes: (1) exact dataset release identifiers for Weibo21, PHEME, and CHEF; (2) explicit confirmation that Weibo21 uses the processed real_cascade version with timestamp-level repost/comment relations; (3) event-disjoint split identifiers; (4) propagation graph extraction scripts and configuration files; (5) lifecycle stage-label generation code and threshold-perturbation settings; (6) random seeds, hyper-parameter files, and model checkpoints; (7) per-seed result tables for main (Table 10), early-window (Table 9), and diagnostic-control (Table 8) experiments, included in the manuscript; seed-wise CSV files and raw run logs archived; (8) raw run logs behind all mean/std tables; (9) QPS serving-test configuration, warm-up policy, batch policy, event-stream input logs, raw latency distributions, and GPU utilization traces; (10) cache-hit assumptions, feature-store configuration, and lightweight serving settings; and (11) verified dataset licenses, code links, and references. The pure propagation-statistics baseline, the protocol-matched GACL re-run, the protocol-adapted 2025-2026 baselines (BCMPL, TSHCL, and PreERD), and the per-seed result tables are included in the manuscript, and the PHEME statistics follow the nine-event release.

Appendix B. Cascade reconstruction sensitivity

The Weibo21 real_cascade version is constructed from samples in the Weibo21 benchmark [26] using our own HDBSCAN-based event-clustering procedure over Chinese-RoBERTa embedding cosine

similarity with a temporal proximity constraint (posts within 48 h). Clustering uses no veracity labels; clusters found after the fact to mix fake and real samples are discarded, so every retained cluster is label-pure by construction (formally, purity, defined as the maximum within-cluster class share, equals 1.0 for all retained clusters), and the sensitivity analysis below is computed on this retained, label-pure event inventory. The audit counts for the main configuration (θ = 0.75) are as follows: clustering produced 764 candidate clusters, of which 74 (9.7%) mixed fake and real samples and were discarded together with their 512 member samples, leaving 690 label-pure clusters covering 8,616 of the 9,128 samples; the fake-to-real ratio is 49.2:50.8 before filtering and 48.7:51.3 after filtering, a shift of less than one percentage point, and the cluster-size distributions before and after filtering will be reported with the released materials. Two hyperparameters control this process: the cosine similarity threshold (θ) and the 48-hour temporal window. Table B1 reports sensitivity to θ to show that the main Early-F1 trend is not an artifact of a single clustering cutoff. A separate label-free sensitivity variant retains all 764 candidate clusters (9,128 samples) without any label-based filtering; its results are reported in Table B2. Across the five seeds, Early-F1@6h moves from 80.1 +/- 1.0 to 78.9 +/- 0.9 and Full F1 from 91.0 +/- 0.4 to 90.3 +/- 0.5 (seed 42: 81.0 to 79.4 and 91.5 to 90.6), so the main Early-F1 trend does not depend on the label-conditioned filtering step. PreERD, the same-stage control, and the propagation-only GBDT degrade by at most 1.1 Early-F1@6h points on the same label-free inventory, so the relative ordering of all compared methods is unchanged.

Table B1. HDBSCAN cosine threshold sensitivity on Weibo21 (seed 42).

Cosine threshold θ	**Event clusters**	**Isolated posts (%)**	**Early-F1@6h**	**Full F1**
0.65	612	8.2	79.5	90.5
0.70	648	5.8	79.8	90.8
0.75 (main)	690	3.5	81.0	91.5
0.80	738	2.1	79.8	90.6
0.85	812	1.0	78.5	89.8

The event cluster count varies from 612 ($\theta = 0.65$, loose clustering) to 812 ($\theta = 0.85$, tight clustering), a 33% range. Early-F1@6h remains within 2.5 points (78.5-81.0) across the full range, with the highest value at the main threshold $\theta = 0.75$. The mild drop at $\theta = 0.85$ suggests that overly tight clustering splits true events into fragments, degrading lifecycle conditioning. The stable ranking across $\theta = 0.65$-0.80 supports the robustness of the main result to event-clustering hyperparameter choice.

Table B2. Label-free event-inventory sensitivity on Weibo21.

Method	**Early-F1@6h (filtered)**	**Full F1 (filtered)**	**Early-F1@6h (label-free)**	**Full F1 (label-free)**
CIML-Net (main)	80.1 +/- 1.0	91.0 +/- 0.4	78.9 +/- 0.9	90.3 +/- 0.5
PreERD	77.7 +/- 0.9	89.5 +/- 0.5	76.6 +/- 1.0	88.6 +/- 0.6
Same-stage control	77.9 +/- 0.9	89.9 +/- 0.5	76.8 +/- 1.0	88.8 +/- 0.6
Propagation-	74.9 +/- 0.6	86.2 +/- 0.5	74.2 +/- 0.7	85.6 +/- 0.6

Method	Early-F1@6h (filtered)	Full F1 (filtered)	Early-F1@6h (label-free)	Full F1 (label-free)
only GBDT				

Note. Results are five-seed mean +/- std. The label-free columns retain all 764 candidate clusters, including 74 label-mixed clusters, without veracity-based filtering. Seed-42 CIML-Net values are 81.0/91.5 (filtered) and 79.4/90.6 (label-free).

Appendix C. CHEF text-only transfer validation

CHEF is an evidence-based Chinese fact-checking dataset with 10,000 claims and evidence [33]; it has no propagation graph or image modality. Because lifecycle, graph, and contrastive modules are inactive, CHEF serves only as a text-encoder transfer validation. Table C1 reports the result (CIML-Net 75.5 Acc, 75.3 F1), which reflects encoder transfer rather than evidence for lifecycle modeling.

Table C1. CHEF text-only transfer validation results.

Method	CHEF Acc	CHEF F1
BERT/RoBERTa [27,28]	72.5 +/- 0.6	72.3 +/- 0.7
BMR	75.2 +/- 0.5	75.0 +/- 0.6
CIML-Net	75.5 +/- 0.5	75.3 +/- 0.6

Appendix D. Baseline reimplementation and adaptation details

This appendix documents how the three recent journal baselines (BCMPL, TSHCL, and PreERD) and GACL are adapted to the common protocol of Section 5.2. For each method we record the code source, the input adaptations required by Weibo21 and PHEME, the treatment of unavailable modalities, the validation search space, the point at which early-window truncation enters the pipeline, and whether encoders are retrained. Adaptations are input-side only: each method's learning objective, loss weights, and architecture remain those of the original paper, and no baseline observes post-cutoff content. Early-window truncation is applied at the data-loader stage, before feature extraction, so truncation cannot leak future information through any preprocessing path. All trainable encoders and downstream modules are optimized under each method's own objective on our training split; CLIP visual encoders are kept frozen where specified by the corresponding configuration, and no checkpoint tuned on Weibo21 or PHEME is imported. Table D1 reports the measured trainable-parameter counts of the adapted models; frozen CLIP branches are excluded from the count.

BCMPL [14] (official authors' release, retrieved January 2026) is designed for balanced multimodal interaction. On Weibo21 it receives the event-slice text and the representative image; it is not evaluated on PHEME, where representative images are too sparse for its image branch (Table 3). Retuned per seed on the validation split: learning rate in {1e-5, 2e-5, 5e-5}, batch size in {16, 32}, dropout in {0.1, 0.3}; all other settings follow the paper defaults.

TSHCL [19] (official authors' release, retrieved January 2026) is a temporal-spatial contrastive method without an image input. On Weibo21 its spatial channel is built from the user-interaction fields of the recollected cascades; on PHEME, from the conversation-thread structure. Search space: learning rate in {1e-5, 2e-5, 5e-5}, contrastive temperature in {0.07, 0.1, 0.2}.

PreERD [18] (official authors' release, retrieved January 2026) predicts the not-yet-formed propagation graph. Under our protocol its predictor is trained on prefix cascades only, and at evaluation time the predicted graph substitutes for the unobserved suffix at each cutoff, exactly as in the original paper. Search space: learning rate in {5e-6, 1e-5, 2e-5}, predictor depth in {1, 2}.

GACL [17] is re-run from the official release with no architectural change; both augmented graph views are constructed from the truncated cascade available at each observation window.

Table D1. Measured parameter counts of the adapted baseline models under the common protocol.

Method	**Encoders in the adapted run**	**Trainable params (M)**
BCMPL [14]	RoBERTa-base + CLIP ViT-B/32 (frozen)	128
TSHCL [19]	RoBERTa-base	116
PreERD [18]	RoBERTa-base + GCN graph predictor	133
CIML-Net (reference)	RoBERTa-large + CLIP ViT-B/16 (frozen) + 2-layer GAT	420