ALMMIT: Agent-Guided Active Learning for Multimodal Misinformation Detection on Temporal Graph Data

Anonymous ACL submission

Abstract

Posting information on a social platform is effortless, often just a single click away; yet, fact-checking mechanisms remain limited, and authenticating information is challenging due to the dynamism of misinformation propagation. The existing misinformation detection frameworks over-rely on annotated data and fall short in managing various modalities (Text, Images, conversational structure) that contribute to misinformation evolution. To mitigate this challenge, we propose ALMMIT, an *Active Learning multimodal framework for Misinformation Detection, guided by a Reinforcement Learning (RL) agent*. The RL agent adaptively directs the active learning process, maintaining momentum across rounds and avoiding bias towards the majority class while operating effectively in few-label and semi-supervised settings to overcome data scarcity within an imbalanced dataset. In addition, ALMMIT employs a modality-aware loss to dynamically adjust modality weights during fine-tuning and active learning, and introduces a composite reward mechanism that balances momentum gain with targeted attention to minority classes. Experimental results show that ALMMIT consistently outperforms recently developed state-of-the-art few-label and supervised approaches.

1 Introduction

The widespread dissemination of misinformation on online social platforms threatens social stability and public well-being. This challenge is exacerbated by the limited availability of effective fact-checking systems. As a result, communities can be influenced and manipulated by the propagation of rumors, even when they originate from seemingly authenticated entities. For example, as highlighted by Tasnim et al. (Tasnim et al., 2020), during the COVID-19 pandemic health systems worldwide faced significant challenges while simultaneously confronting a surge of misinformation. Such misinformation undermined appropriate health practices, encouraged harmful behaviors, and accelerated disease transmission, leading to adverse physical and psychological consequences. Consequently, misinformation detection has attracted substantial research attention, particularly with the rise of advanced techniques such as Large Language Models (LLMs).

Early approaches primarily relied on textual features and sentiment analysis (Castillo et al., 2011; Ma et al., 2015; Tolosi et al., 2016; Enayet and El-Beltagy, 2017; Kumar et al., 2022). Although these methods achieved promising results, they often overlooked the temporal dynamics of misinformation propagation. In practice, user interactions such as sharing, commenting, and replying play a crucial role in shaping how misinformation spreads. Ignoring these interaction patterns limits the ability to accurately capture the evolution of misinformation within social networks (Lin et al., 2021; Liu et al., 2022; Li et al., 2023; Hao et al., 2024; Li et al., 2024b).

Beyond temporal and interaction signals, multimodal content such as images and videos also plays an important role in misinformation dissemination. However, despite recent progress (Zhao et al., 2023), visual information remains underutilized in many existing approaches. Recent efforts attempt to address this limitation by combining synthetic and real data (Zeng et al., 2024) and constructing cross-modal graphs through label propagation (Zhao et al., 2023). For example, GAMED introduces a multimodal detection framework with modal decoupling, multi-expert networks, and veto voting (Shen et al., 2025). FCN-LP builds CLIP-based tweet graphs with domain generalization (Zhao et al., 2023), while MMAdapt leverages knowledge graphs with uncertainty-aware adaptation for health misinformation detection (Shang et al., 2024). In addition, (Bartlett et al., 2025) highlights the growing role of short-form videos with misleading captions in spreading misinforma-

tion on social media platforms. Nevertheless, many existing methods still rely heavily on annotated data and often overlook the dynamic contributions of different modalities during misinformation propagation (Thibault et al., 2025).

To address these limitations, we propose **ALMMIT**, an oracle-free adaptive active learning framework with the following contributions:

1. **Stage-aware multimodal segmentation:** ALMMIT introduces temporal variable-length stage partitioning while aligning text, image, and graph modalities within conversation threads.

2. **Modality-aware loss with presence gating:** a learnable objective that dynamically reweights modalities and remains robust to missing inputs, addressing a key limitation of prior multimodal approaches.

3. **Dual-adaptive active learning:** a reinforcement learning controller that jointly adjusts acquisition weights (α) and modality weights (β), unifying sample selection and modality fusion.

4. **Composite reward for fairness:** a reinforcement-driven reward that balances learning momentum with minority-class rarity to improve stability under class imbalance.

2 Related Work

2.1 Misinformation Propagation

Disinformation on social media often follows identifiable and recognizable propagation patterns and attributes. These circumstances have led researchers to analyze the problem from three main perspectives: sequence, structure, and stage. Early works have revolved around sequential dynamics, focusing on how messages evolve and how contextual signals would shift during propagation (Tsallis, 2022; Anand et al., 2014; Schmid Jr, 1947; Zubiaga et al., 2015). While useful, sequential modeling alone cannot capture the complex web of interactions that drive diffusion in real-world networks. To address this limitation, scholars proposed a range of graph-based approaches concerning representation learning and classification (Yu et al., 2019; Wei et al., 2019; Bian et al., 2020; Khoo et al., 2020). These methods leverage graph neural networks to represent structural relationships and incorporate semantic information. However, they often fall short in timely detection and devote limited attention to how interactions evolve. Building on these directions, MsDD (Hao et al., 2024) have segmented conversation thread evolution into variable-length stages using graph entropy to capture more meaningful temporal patterns, while ADAAT (Wang et al., 2025) emphasizes robustness by combining adaptive augmentation, adversarial training, and contrastive learning. Nevertheless, both approaches remain limited to text–graph modalities under full supervision, without handling multimodal posts, adaptive modality weighting across stages, or scenarios with few labeled examples.

2.2 Active Learning

Active learning aims to select the most informative samples to improve model performance while reducing annotation cost, initially relying on uncertainty-based criteria such as entropy (Ren et al., 2021), EMCM (Cai et al., 2013), and Bayesian approaches including BALD and its extensions (Gal et al., 2017; Kirsch et al., 2019; Tran et al., 2019). Subsequent studies emphasized consistency (Gao et al., 2020), diversity and distributional similarity (Margatina et al., 2021; Sener and Savarese, 2017), VAE-based selection (Sinha et al., 2019), and hybrid strategies that combine informativeness and representativeness (Zhang et al., 2022a; Zhdanov, 2019; Ash et al., 2019), with extensions to multimodal settings (Shen et al., 2023; Li et al., 2024a). In graph-based active learning, methods exploiting entropy, density, structural signals, and reinforcement learning have been proposed (Cai et al., 2017; Gao et al., 2018; Chen et al., 2019; Zhang et al., 2021a,b; Hu et al., 2020; Cui et al., 2022; Zhang et al., 2022b), while class imbalance is partially mitigated by GraphCBAL (Yu et al., 2024); however, these approaches largely overlook modality-specific contributions within threads and lack adaptive weighting of acquisition components.

3 Methodology

3.1 Our Novel ALMMIT Model

ALMMIT (Figure 1) is a dual-adaptive framework for misinformation detection in conversation threads. Remarkable Sub-Graph Selectors (RSGS) extract stage-aware subgraphs that preserve temporal propagation while integrating text, image, and structural signals. These subgraphs are pro-

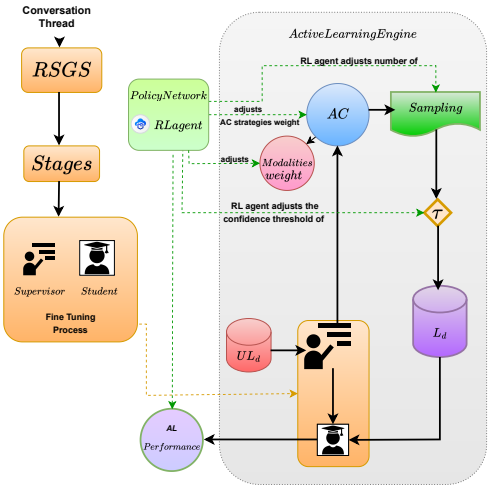


Figure 1: **Abstract ALMMIT diagram**: Student and Supervisor have been fine-tuned, and then these fine-tuned models have been used in the active learning engine.

cessed by a Supervisor trained with limited annotations. An Active Learning Engine then expands the training set by scoring unlabeled samples using uncertainty, diversity, representativeness, and curriculum signals. A reinforcement learning controller dynamically adjusts acquisition and modality weights during training, while pseudo-labels further refine stage-level weighting and acquisition strategies. This design enables adaptation both within stages (via modality fusion) and across active learning rounds, allowing ALMMIT to perform effectively under limited supervision.

3.2 Notations and Problem Description

Let $\mathcal{C} = (\mathcal{T}, \mathcal{I}, \mathcal{G})$ denote a conversation thread, where $\mathcal{T} = \{t_1, \ldots, t_n\}$ is the set of textual posts, $\mathcal{I} = \{i_1, \ldots, i_n\}$ are the associated images (if present), and $\mathcal{G} = (V, E)$ is the interaction graph with V the set of posts/users and E the reply or retweet relations. Each thread $\mathcal{C}$ is associated with a binary label $y \in \{0, 1\}$ indicating whether it propagates misinformation. We further assume that $\mathcal{G}$ can be segmented into K temporal *stages* $\{S_1, \ldots, S_K\}$, where each stage contains a subset of posts and their corresponding interactions.

The training data consists of a small labeled subset $L = \{(\mathcal{C}_j, y_j)\}_{j=1}^{\ell}$ and a large unlabeled pool $U = \{\mathcal{C}_j\}_{j=\ell+1}^{N}$, with $\ell \ll N$. The task is to learn a classifier

$$f : (\mathcal{T}, \mathcal{I}, \mathcal{G}) \mapsto y$$

that generalizes to unseen threads while operating under annotation scarcity. This setting requires the learner to jointly address three challenges: (i) adaptively weight modality contributions (text, image, and graph) across stages, (ii) capture dynamic interaction evolution through stage segmentation, and (iii) actively select or pseudo-label samples from U under a strict budget. To formalize adaptivity, let $\beta^{(k)} = (\beta_T^{(k)}, \beta_I^{(k)}, \beta_G^{(k)})$ denote the modality weights assigned at stage S_k, and let $\alpha^{(t)} = (\alpha_u^{(t)}, \alpha_r^{(t)}, \alpha_d^{(t)}, \alpha_c^{(t)})$ denote the acquisition weights for uncertainty, representativeness, diversity, and curriculum at active learning round t. The optimization objective is to learn parameters θ of f such that performance on a held-out test set T_{test} is maximized while respecting limited supervision and adaptivity constraints:

$$\theta^* = \arg\max_{\theta} \; E_{(\mathcal{C},y)\in T_{test}} \, 1\Big[f_\theta(\mathcal{C}; \alpha^{(1..K)}, \beta^{(1..R)}) = y\Big] \,,$$

where R is the number of AL rounds. This definition not only extends stage segmentation to multimodal alignment (text, image, and graph) but also captures the dual adaptivity of our setting: modality weights $\beta^{(k)}$ vary with stage evolution, while acquisition weights $\alpha^{(t)}$ evolve across AL iterations to balance competing selection objectives.

3.3 ALMMIT Description

3.3.1 Remarkable Sub-Graph Selectors (RSGS)

RSGS decomposes a conversation thread into *variable-length, stage-level sub-graphs* that retain both structural and semantic evolution. Let the thread interaction graph be $\mathcal{G} = (V, E)$ with per-node timestamps $\{t_v\}_{v\in V}$. For any time u, we define the cumulative subgraph

$$\begin{aligned} G(u) &= \big(V(u), E(u)\big), \\ V(u) &= \{v \in V : t_v \le u\}, \\ E(u) &= \{(i,j) \in E : \max(t_i, t_j) \le u\}. \end{aligned} \tag{1}$$

Let $d_i(u)$ be the degree of node i in $G(u)$ and $D(u) = \sum_{i\in V(u)} d_i(u)$. We measure structural complexity using the degree entropy:

$$H(u) = -\sum_{i\in V(u)} \frac{d_i(u)}{D(u)} \ln\left(\frac{d_i(u)}{D(u)}\right). \tag{2}$$

Starting from a stage start time s, RSGS advances a *cumulative window* in steps of length τ, i.e., $u_k = s + k\tau$ for $k = 1, 2, \ldots$. At each step it computes the entropy increment

$$\Delta H_k = H(u_k) - H(u_{k-1}), \tag{3}$$

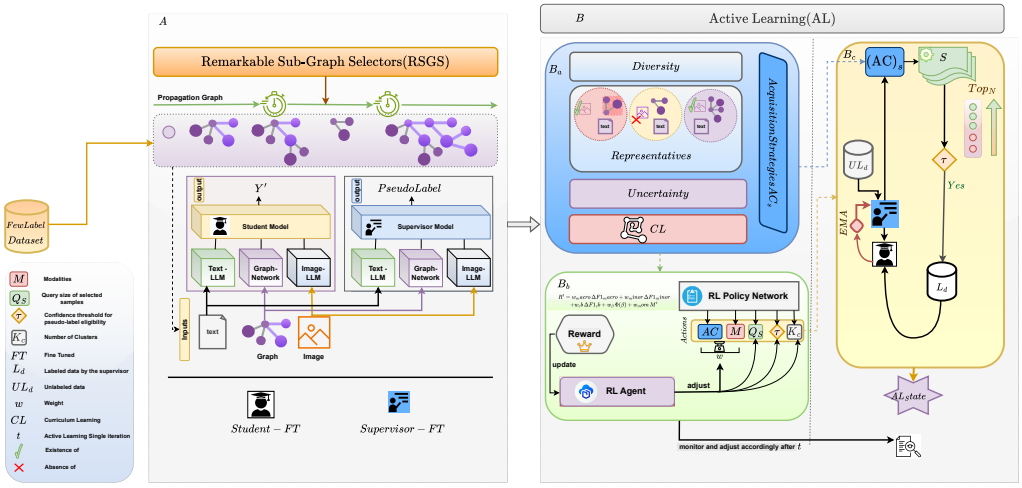


Figure 2: **ALMMIT**: Agent-Guided Active Learning for Multimodal Misinformation Detection. In Phase A, both the Student and Supervisor are fine-tuned. In Phase B, Active Learning leverages these fine-tuned models. Consistent coloring illustrates the connections between phases.

normalizes by τ, and evaluates a *tangent-like contrast* between consecutive normalized changes:

$$\tan\theta_k = \frac{\left|\frac{\Delta H_k}{\tau} - \frac{\Delta H_{k-1}}{\tau}\right|}{1 + \left(\frac{\Delta H_k}{\tau}\right)\left(\frac{\Delta H_{k-1}}{\tau}\right)}. \quad (4)$$

A *stage boundary* is declared at u_k if $\tan\theta_k > \mu$, and the stage sub-graph S is instantiated as the nodes and edges whose timestamps fall in $[s, u_k]$. If no boundary is detected, the remainder $[s, t_{\max}]$ forms the final stage.

Each detected stage inherits all modalities (text, image, graph) active within its temporal span, yielding a compact and temporally coherent *multimodal stage unit*. These stage units are then passed downstream for adaptive fusion and active learning. The fusion module reweights modality contributions per stage via $\beta^{(k)}$, while the active learning engine iteratively updates acquisition weights across rounds via $\alpha^{(t)}$.

3.3.2 Supervisor Fine-Tuning

The second stage of our framework is a multimodal *Supervisor* that learns to integrate text, image, and graph signals into a unified prediction. Each conversation thread is represented as a sequence of posts enriched with visual features and structured by stage-level sub-graphs. The Supervisor employs three modality-specific encoders: a transformer-based language model for textual embeddings, a vision backbone for image features, and a dynamic graph neural network for structural representations. These three embeddings are concatenated and mapped to logits by a lightweight fusion head:

$$\mathbf{z} = \big[f_T(\mathcal{T}) \parallel f_I(\mathcal{I}) \parallel f_G(\mathcal{G})\big], \qquad \boldsymbol{\ell} = \mathrm{MLP}(\mathbf{z}), \quad (5)$$

where f_T, f_I, and f_G denote the modality encoders, $\mathbf{z}$ is the fused embedding, and $\boldsymbol{\ell}$ the output logits.

A critical design choice is that the Supervisor is *fine-tuned under few-label supervision*. Instead of relying on large fully annotated corpora, it is trained with only a small fraction of labeled threads (e.g., 15% of the dataset), reflecting realistic annotation scarcity in misinformation detection. This few-label fine-tuning not only makes the Supervisor robust in low-resource conditions but also primes it to provide reliable pseudo-labels and teacher guidance during the subsequent active learning process.

The primary training objective is a class-weighted cross-entropy loss that compensates for label imbalance by assigning larger weights to under-represented classes:

$$\mathcal{L}_{\mathrm{CE}} = -\frac{1}{B}\sum_{b=1}^{B}\sum_{c=1}^{C} w_c\, 1[y_b = c]\, \log \hat{p}_{b,c}, \quad (6)$$

where w_c denotes the class weights and $\hat{p}_{b,c}$ the predicted probability for class c of sample b. A distinctive feature of our Supervisor is the introduction of a *modality-aware loss*, which explicitly distinguishes between samples that are text-only, samples with visual content, and the conversational graph. Let $\mathcal{L}_{\mathrm{T}}$, $\mathcal{L}_{\mathrm{V}}$, and $\mathcal{L}_{\mathrm{G}}$ denote the class-weighted cross-entropy losses for each group. Instead of manually assigning weights, we introduce learnable parameters $\theta = (\theta_{\mathrm{T}}, \theta_{\mathrm{V}}, \theta_{\mathrm{G}})$ and

normalize them via softplus:

$$w_k = \frac{\mathrm{softplus}(\theta_k)}{\sum_{j\in\{\mathrm{T,V,G}\}} \mathrm{softplus}(\theta_j)}, \qquad k \in \{\mathrm{T,V,G}\}. \tag{7}$$

To ensure that absent modalities (e.g., no images in the batch) do not contribute, we apply *presence gating*. Define $\pi_k = 1$ if $\mathcal{L}_k$ is computable in the current batch and $\pi_k = 0$ otherwise. The weights are then re-normalized over the present groups:

$$\tilde{w}_k = \frac{\pi_k\, w_k}{\sum_{j\in\{\mathrm{T,V,G}\}} \pi_j\, w_j}, \qquad \mathcal{L}_{\text{mod}} = \sum_{k\in\{\mathrm{T,V,G}\}} \tilde{w}_k\, \mathcal{L}_k. \tag{8}$$

This design enables the Supervisor to adaptively learn how to balance supervision from threads with and without images, preventing bias toward vision-rich samples and ensuring robustness in settings where images are scarce. The modality-aware loss is therefore not only a stabilizing mechanism but also a novel contribution of our framework, supporting the broader theme of adaptivity. In summary, the Supervisor serves as a strong teacher model *trained under few-label conditions*, producing high-quality multimodal representations and pseudo-labels that directly fuel the dual-adaptive active learning engine. Unlike existing multimodal misinformation detection methods that assume full supervision, our Supervisor is deliberately designed for few-label fine-tuning, enabling practical deployment in real-world low-annotation scenarios.

3.3.3 B Active Learning Engine

In the final stage of our framework, we employ a dedicated Active Learning (AL) engine that connects the supervisor to the unlabeled pool. The engine determines which samples to query or pseudo-label, how to balance acquisition criteria and modality contributions, and how to refine the student under limited supervision. It is structured into three parts: **$\mathbf{B_a}$) Acquisition Strategies ($\mathbf{AC_s}$)**, which score and filter candidate samples, **$\mathbf{B_b}$) Reinforcement Learning Policy**, which adaptively adjusts acquisition and modality weights, and **$\mathbf{B_c}$) Active Learning Training**, which integrates pseudo-labels, optimizes the student, and updates the teacher via EMA.

3.3.4 $\mathbf{B_a}$) Acquisition Strategies ($\mathbf{AC_s}$)

The acquisition module is responsible for selecting the most valuable conversation threads from the unlabeled pool U to be queried or pseudo-labeled. To balance informativeness and coverage, we integrate four complementary criteria: *uncertainty*, *representativeness*, *curriculum*, and a lightweight *diversity guard*. Each criterion produces a score, which is combined adaptively using policy-controlled weights.

Uncertainty. We estimate predictive uncertainty via the entropy of the calibrated teacher distribution:

$$s_{\text{unc}}(x) = -\sum_{c=1}^{C} p(c|x)\, \log p(c|x), \tag{9}$$

where $p(c|x)$ is the softmax probability for class c after temperature calibration.

Curriculum. To promote stable training, we incorporate a curriculum signal that prioritizes confident (easy) samples in early rounds and gradually shifts emphasis toward harder ones as learning progresses:

$$s_{\text{cur}}(x) = \gamma_t \cdot \max_c\, p(c|x), \tag{10}$$

where $\gamma_t = \min(1, t/T_{\text{warm}})$ is a warm-up factor depending on round t and curriculum horizon T_{warm}.

Representativeness. To ensure that queried samples reflect the global pool distribution, we compute representativeness by clustering embeddings $\mathbf{h}(x)$ of pool samples using KMeans or a Bayesian Gaussian mixture. For each item x, the score is inversely related to its squared distance from the cluster center:

$$s_{\text{rep}}(x) = 1 - \frac{\|\mathbf{h}(x) - \mu_{c(x)}\|^2}{d_{\max}}, \tag{11}$$

where $\mu_{c(x)}$ is the assigned cluster centroid and $d_{\max}$ the maximum observed distance for normalization, ensuring scores fall in a bounded range.

Diversity. While representativeness prevents the system from focusing on outliers, it does not guarantee diversity in the selected batch. To avoid redundancy, we enforce a diversity guard by allocating quotas across clusters and classes, ensuring that each batch covers multiple regions of the pool space. This cluster-quota mechanism provides implicit diversity without adding an additional scoring dimension.

Final Score. For each sample x, the overall acquisition score is a weighted combination of the three principal signals:

$$s(x) = \alpha_{\text{cur}} s_{\text{cur}}(x) + \alpha_{\text{rep}} s_{\text{rep}}(x) + \alpha_{\text{unc}} s_{\text{unc}}(x), \tag{12}$$

where $\alpha = (\alpha_{\text{cur}}, \alpha_{\text{rep}}, \alpha_{\text{unc}})$ are policy-controlled weights that evolve across rounds (see Sec. B_b). Diversity guards are applied after scoring to regulate cluster balance in the selected batch.

This acquisition design not only combines informativeness and representativeness but also ensures stability (via curriculum) and coverage (via diversity guard). Unlike prior graph-based active learning methods that emphasize single criteria (e.g., uncertainty or centrality), our approach introduces a *dual-adaptive mechanism*: acquisition weights α are dynamically adjusted by the RL agent, while modality weights β (Sec. B_c) are concurrently optimized to handle multimodal imbalance.

3.3.5 $\mathbf{B_b}$) Reinforcement Learning (RL) Policy Network and Agent

Motivation for RL Control. The active learning process is inherently dynamic: as the labeled set grows across rounds, the relative importance of acquisition strategies (e.g., uncertainty, representativeness, and curriculum) and modality contributions may shift. Static weighting schemes assume fixed importance for these components and therefore cannot adapt to evolving data distributions. To address this, we employ a reinforcement learning (RL) controller that learns to adjust acquisition and modality weights jointly across active learning rounds, enabling the system to balance exploration, representativeness, and modality reliability in a data-driven manner.

While acquisition scores provide candidate utilities, their relative importance changes across rounds. To adaptively control this balance, we introduce a reinforcement learning (RL) agent that dynamically tunes acquisition weights and other hyperparameters of the AL loop. The agent follows a Gaussian REINFORCE framework.

Policy Network. Let the environment state at round t be represented by a feature vector $\mathbf{s}^{(t)} \in R^{10}$ that encodes progress and model status (e.g., budget remaining, class entropy, recent validation performance). A shared trunk projects $\mathbf{s}^{(t)}$ into a hidden representation $h^{(t)}$, which is mapped into action heads. Each head outputs a Gaussian distribution with mean μ and log-variance $\log \sigma$:

$$\mu, \log \sigma = W h^{(t)} + b, \quad a \sim \mathcal{N}(\mu, \sigma^2). \tag{13}$$

The policy controls five sets of actions:

- $\alpha \in \Delta^2$: simplex weights over acquisition criteria $\{s_{\text{cur}}, s_{\text{rep}}, s_{\text{unc}}\}$,
- $\beta \in \Delta^2$: modality weights (text, image, graph) in the modality-aware loss,
- $q \in [q_{\min}, q_{\max}]$: query fraction of the pool to be selected,
- $\tau \in [\tau_{\min}, \tau_{\max}]$: confidence threshold for pseudo-label acceptance,
- $K \in [K_{\min}, K_{\max}]$: number of clusters for representativeness/diversity balancing.

Learning. At each round, the agent samples actions $a^{(t)}$ from the policy and receives a reward $R^{(t)}$ derived from validation performance improvements (macro-F1, minority-F1, class balance) and modality-balance shaping. The policy parameters are updated using the REINFORCE gradient:

$$\nabla_\theta J(\theta) = E\Big[(R^{(t)} - b^{(t)})\, \nabla_\theta \log \pi_\theta(a^{(t)} | \mathbf{s}^{(t)}) + \lambda H(\pi_\theta)\Big], \tag{14}$$

where $b^{(t)}$ is a moving-average baseline, and $H(\pi_\theta)$ is an entropy regularizer encouraging exploration.

Reward Shaping. Instead of using raw accuracy gains, the reward $R^{(t)}$ integrates multiple signals to ensure stable and balanced progress. It combines improvements in macro- and minority-class F1, class-balance metrics, a modality-balance bonus derived from loss variance, and a momentum term computed from exponential moving averages of recent validation trends:

$$\begin{aligned} R^{(t)} = \; & w_{\text{macro}}\, \Delta\text{F1}_{\text{macro}} + w_{\text{minor}}\, \Delta\text{F1}_{\text{minor}} \\ & + w_{\text{cb}}\, \Delta\text{F1}_{\text{cb}} + w_\beta\, \Phi(\beta) + w_{\text{mom}}\, M^{(t)}. \end{aligned} \tag{15}$$

Here $\Phi(\beta)$ penalizes imbalance among modality weights, while $M^{(t)}$ rewards sustained positive momentum. Rewards are clipped to a bounded range and normalized by an EMA baseline, which reduces variance in policy updates.

Stability Aids. To prevent collapse, we enforce action constraints: (i) q is bounded below by an annealed floor $q_{\text{floor}}(t)$, (ii) α may be reset to uniform when a round yields highly negative rewards,

and (iii) β is updated using an exponential moving average (EMA) to smooth fluctuations. These safeguards ensure that the RL controller remains stable even under noisy pseudo-label training.

Overall, the RL module empowers the engine to *adaptively re-weight both acquisition signals and modality contributions across AL rounds*, a dual-adaptive feature absent in prior AL frameworks.

3.3.6 B_c) Active Learning Training

Once candidate samples are selected, the training phase integrates pseudo-labels generated by the teacher. To mitigate noise, we adopt strict eligibility checks based on confidence, class-proportion gating, and margin thresholds. Let $\hat{y}$ denote the pseudo-label and $p(c|x)$ the calibrated teacher probability. A sample x is eligible if

$$p(\hat{y}|x) \geq \tau, \quad (p_1 - p_2) \geq \delta, \quad \rho(\hat{y}) < \rho_{\max},$$

where $p_1 - p_2$ is the top-2 margin, δ a margin floor, and $\rho(\hat{y})$ the relative class proportion.

Loss formulation. The student is trained using a combination of class-balanced cross-entropy (CE), modality-aware CE, and KL divergence. First, the class-balanced CE compensates for skewed label distributions:

$$\mathcal{L}_{\text{CE}} = -\sum_i w_{y_i} \log p_\theta(y_i|x_i), \tag{16}$$

where w_{y_i} is the effective weight derived from inverse class frequency. Second, the modality-aware CE computes class-balanced losses individually for text-only ($\mathcal{L}_T$), image-present ($\mathcal{L}_I$), and graph ($\mathcal{L}_G$) samples, and recombines them using learnable weights β:

$$\mathcal{L}_{\text{mod}} = \tilde{\beta}_T \mathcal{L}_T + \tilde{\beta}_I \mathcal{L}_I + \tilde{\beta}_G \mathcal{L}_G, \tag{17}$$

where $\beta = \text{softmax}(\theta)$ ensures non-negativity and normalization. To handle absent modalities, we apply *presence gating*: if $\mathcal{L}_k$ is undefined in the current batch, the corresponding weight is set to zero and the remaining weights are renormalized, i.e.,

$$\tilde{\beta}_k = \frac{\pi_k \beta_k}{\sum_{j \in \{\text{T,I,G}\}} \pi_j \beta_j}, \qquad \pi_k \in \{0, 1\}, \tag{18}$$

so that $\sum_k \tilde{\beta}_k = 1$ holds over the modalities present. Finally, we impose KL consistency between student and teacher distributions:

$$\mathcal{L}_{\text{KL}} = \frac{1}{N} \sum_{i=1}^{N} \text{KL}\big(p_T(\cdot|x_i) \,\|\, p_S(\cdot|x_i)\big), \tag{19}$$

with annealed weight $\lambda_{\text{KL}}(t)$ that increases over rounds to stabilize pseudo-label guidance.

Overall objective. The student optimization at round t is:

$$\mathcal{L}^{(t)} = \mathcal{L}_{\text{CE}} + \mathcal{L}_{\text{mod}} + \lambda_{\text{KL}}(t)\, \mathcal{L}_{\text{KL}}. \tag{20}$$

Note that both $\mathcal{L}$CE and $\mathcal{L}$mod are computed against pseudo-labels provided by the teacher, ensuring that training remains fully oracle-free.

Teacher update. To avoid teacher drift, we adopt an exponential moving average (EMA) update:

$$\theta_T \leftarrow \tau\, \theta_T + (1 - \tau)\, \theta_S, \tag{21}$$

where θ_T and θ_S are teacher and student parameters, and τ is a decay rate that anneals with the round index. This gradually improves the teacher while preserving stability.

4 Experiments and Results

To evaluate the effectiveness of ALMMIT in misinformation detection, we conducted a comprehensive set of experiments.

4.1 Datasets

We have utilized the Pheme (Zubiaga et al., 2016) dataset that comprises conversation threads from X (formerly Twitter), centered on five breaking news events. Each event includes a set of sources (The main post and its interactions), including a large number of texts and images. We have considered features including conversation threads, texts, images, and temporal data. To demonstrate flexibility, we also incorporated the Twitter15 (Ma et al., 2017) dataset, which provides only text and conversational structure without images. Detailed statistical characteristics are reported in Appendix 8.

4.2 Discussion and result

All evaluation metrics follow standard definitions. Accuracy is computed at the thread level, while Precision, Recall, and F1 scores in Table 1 are macro-averaged unless otherwise specified. Table 1 compares ALMMIT with recent few-label and semi-supervised baselines on the Pheme and Twitter15 datasets. The results show that ALMMIT consistently achieves stronger overall performance, demonstrating the effectiveness of combining multimodal modeling with adaptive active learning under limited supervision. Concerning

Table 1: Comparison between ALMMIT and baselines on Pheme and Twitter15 datasets. The best results are highlighted in bold.

Method	Acc	Real News			Fake News		
		P	R	F1	P	R	F1
Pheme dataset							
FLAL	79.4	83.3	76.7	80.1	81.7	71.9	76.7
IdoFew	73.8	80.7	62.9	70.7	69.5	84.9	76.4
BERTIT:Cluster	76.0	83.1	65.4	73.2	71.4	86.6	78.3
ALMMIT	**82.64**	**91.38**	**81.35**	**86.08**	**70.19**	**85.13**	**76.94**
Twitter15							
FLAL	80.2	81.1	19.0	27.7	74.0	30.1	48.5
IdoFew	74.9	56.2	13.4	21.6	76.1	**96.4**	85.1
BERTIT:Cluster	74.4	51.5	19.0	27.7	76.8	93.8	84.5
ALMMIT	**82.35**	**65.625**	**64.61**	**65.116**	**87.958**	88.42	**88.18**

Table 2: Comparison with fully supervised text-based approaches on the Pheme dataset. **ALMMIT* is under the Semi-supervised setting**. The best results are highlighted in bold.

Method	**Accuracy**	**F1**
ACAMI	75.78	61.66
HGCN-RNN	85.28	81.18
Bi-GCN	84.97	80.32
StA-PLAN	83.36	79.78
RDLNP	84.31	80.78
DynGCN	86.78	82.40
ALMMIT*	**86.89**	**84.07**

Table 3: Comparison with multimodal baselines on the Pheme dataset. **ALMMIT* is under the semi-supervised setting**. All Precision, Recall, and F1 scores are macro-averaged. The best results are highlighted in bold.

Method	**Accuracy**	**Precision**	**Recall**	**F1**
EANN	70.17	71.28	67.36	69.10
SpotFake	81.37	79.53	81.22	79.43
MVAE	77.83	73.82	73.45	72.21
SAFE	81.25	79.92	79.11	79.69
MCAN	80.74	79.21	79.64	80.15
HMCAN	86.36	83.18	83.81	83.45
CLIP	80.36	84.43	89.12	86.71
CLIP + FCN-LP	84.65	**88.79**	**89.81**	**89.30**
ALMMIT*	**86.89**	89.25	81.81	84.08

the Pheme dataset, ALMMIT achieves an accuracy of 82.64% and an F1 score of 84.07, outperforming leading few-label baselines. It obtains a recall of 85.13% for the *Fake News* class and a Real News F1 of 86.08, surpassing FLAL, IdoFew, and BERTIT:Cluster across both global and class-level metrics. High recall in the misinformation class reduces false negatives and highlights the effectiveness of the proposed stage-aware design and modality-aware active learning strategy. On Twitter15, ALMMIT achieves 82.35% accuracy, outperforming FLAL (80.2%), IdoFew (74.9%), and BERTIT:Cluster (74.4%). It also records a Fake News F1 of 65.12 and a Real News F1 of 88.18, indicating balanced performance across both classes. In contrast, several baselines show stronger performance on one class while degrading on the other. Despite being trained under semi-supervised conditions, ALMMIT matches or surpasses fully supervised text-based baselines on Pheme (Table 2), including Bi-GCN, RDLNP, and DynGCN, demonstrating strong label efficiency. Compared with recent multimodal approaches (Table 3), ALMMIT achieves the highest overall accuracy (86.89%) while remaining competitive in F1 despite CLIP+FCN-LP attaining higher precision and F1. These results confirm the robustness and stability of ALMMIT across evaluation metrics. An ablation study is provided in Appendix B.

5 Conclusion

We presented ALMMIT, a novel framework for misinformation detection that achieves strong performance under few-label and semi-supervised settings. By introducing a modality-aware loss and a reinforcement learning–driven adaptive active learning strategy that dynamically adjusts modality weights across propagation stages, ALMMIT effectively addresses limitations of existing multimodal approaches. Extensive experiments on the Pheme and Twitter15 datasets demonstrate that ALMMIT outperforms several recent state-of-the-art methods.

6 Limitations

While the proposed framework demonstrates strong performance under few-label and semi-supervised settings, several directions remain for further improvement. First, although ALMMIT effectively integrates multimodal signals and adaptive active learning, incorporating a more explicit explainability layer could help provide clearer insights into how different modalities and structural interactions contribute to the final prediction. Such transparency would further enhance the interpretability of misinformation detection systems. Second, the current evaluation focuses on two widely used benchmark datasets. Expanding the evaluation to additional datasets with more diverse topics and media formats would further validate the generalizability of the framework. Finally, extending the framework to incorporate additional modalities and exploring the behavior of LLM-based components under more extreme zero-shot conditions would provide valuable insights.

References

Abdulrahman Alalawi, Abdullah Alsuhaibani, Usman Naseem, Basem Suleiman, Shoaib Jameel, and Imran Razzak. 2025. Few labels with active learning: From weak to strong labels for misinformation detection. In *Companion Proceedings of the ACM on Web Conference 2025*, pages 2592–2596.

Abdullah Alsuhaibani, Hamad Zogan, Imran Razzak, Shoaib Jameel, and Guandong Xu. 2024. Idofew: intermediate training using dual-clustering in language models for few labels text classification. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, pages 18–27.

Kartik Anand, Dmitri Krioukov, and Ginestra Bianconi. 2014. Entropy distribution and condensation in random networks with a given degree distribution. *Physical Review E*, 89(6):062807.

Jordan T Ash, Chicheng Zhang, Akshay Krishnamurthy, John Langford, and Alekh Agarwal. 2019. Deep batch active learning by diverse, uncertain gradient lower bounds. *arXiv preprint arXiv:1906.03671*.

Andrew Bartlett, Waheeb Yaqub, Basem Suleiman, and Manoranjan Mohanty. 2025. Misinformation in reels, influence of contextual superimposed texts in short videos. In *International Conference on Web Information Systems Engineering*, pages 3–14. Springer.

Tian Bian, Xi Xiao, Tingyang Xu, Peilin Zhao, Wenbing Huang, Yu Rong, and Junzhou Huang. 2020. Rumor detection on social media with bi-directional graph convolutional networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 549–556.

Hongyun Cai, Vincent W Zheng, and Kevin Chen-Chuan Chang. 2017. Active learning for graph embedding. *arXiv preprint arXiv:1705.05085*.

Wenbin Cai, Ya Zhang, and Jun Zhou. 2013. Maximizing expected model change for active learning in regression. In *2013 IEEE 13th international conference on data mining*, pages 51–60. IEEE.

Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web*, pages 675–684.

Xia Chen, Guoxian Yu, Jun Wang, Carlotta Domeniconi, Zhao Li, and Xiangliang Zhang. 2019. Activehne: Active heterogeneous network embedding. *arXiv preprint arXiv:1905.05659*.

Jiho Choi, Taewook Ko, Younhyuk Choi, Hyungho Byun, and Chong-kwon Kim. 2021. Dynamic graph convolutional networks with attention mechanism for rumor detection on social media. *Plos one*, 16(8):e0256039.

Limeng Cui, Xianfeng Tang, Sumeet Katariya, Nikhil Rao, Pallav Agrawal, Karthik Subbian, and Dongwon Lee. 2022. Allie: Active learning on large-scale imbalanced graphs. In *Proceedings of the ACM web conference 2022*, pages 690–698.

Omar Enayet and Samhaa R El-Beltagy. 2017. Niletmrg at semeval-2017 task 8: Determining rumour and veracity support for rumours on twitter. In *Proceedings of the 11th international workshop on semantic evaluation (SemEval-2017)*, pages 470–474.

Yarin Gal, Riashat Islam, and Zoubin Ghahramani. 2017. Deep bayesian active learning with image data. In *International conference on machine learning*, pages 1183–1192. PMLR.

Li Gao, Hong Yang, Chuan Zhou, Jia Wu, Shirui Pan, and Yue Hu. 2018. Active discriminative network representation learning. In *27th International Joint Conference on Artificial Intelligence (IJCAI)*. International Joint Conference on Artificial Intelligence (IJCAI).

Mingfei Gao, Zizhao Zhang, Guo Yu, Sercan Ö Arık, Larry S Davis, and Tomas Pfister. 2020. Consistency-based semi-supervised active learning: Towards minimizing labeling cost. In *European Conference on Computer Vision*, pages 510–526. Springer.

Xiaorong Hao, Bo Liu, Xinyan Yang, Xiangguo Sun, Qing Meng, and Jiuxin Cao. 2024. Multi-stage dynamic disinformation detection with graph entropy guidance. *World Wide Web*, 27(2):8.

Shengding Hu, Zheng Xiong, Meng Qu, Xingdi Yuan, Marc-Alexandre Côté, Zhiyuan Liu, and Jian Tang. 2020. Graph policy network for transferable active learning on graphs. *Advances in Neural Information Processing Systems*, 33:10174–10185.

Dhruv Khattar, Jaipal Singh Goud, Manish Gupta, and Vasudeva Varma. 2019. Mvae: Multimodal variational autoencoder for fake news detection. In *The world wide web conference*, pages 2915–2921.

Ling Min Serena Khoo, Hai Leong Chieu, Zhong Qian, and Jing Jiang. 2020. Interpretable rumor detection in microblogs by attending to user interactions. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 8783–8790.

Andreas Kirsch, Joost Van Amersfoort, and Yarin Gal. 2019. Batchbald: Efficient and diverse batch acquisition for deep bayesian active learning. *Advances in neural information processing systems*, 32.

Akshi Kumar, MPS Bhatia, and Saurabh Raj Sangwan. 2022. Rumour detection using deep learning and filter-wrapper feature selection in benchmark twitter dataset. *Multimedia Tools and Applications*, 81(24):34615–34632.

An Lao, Chongyang Shi, and Yayi Yang. 2021. Rumor detection with field of linear and non-linear propagation. In *Proceedings of the web conference 2021*, pages 3178–3187.

Fangfang Li, Zhi Liu, Junwen Duan, Xingliang Mao, Heyuan Shi, and Shichao Zhang. 2023. Exploiting conversation-branch-tweet hypergraph structure to detect misinformation on social media. *ACM Transactions on Knowledge Discovery from Data*, 18(2):1–20.

Xinyu Li, Wenqing Ye, Yueyi Zhang, and Xiaoyan Sun. 2024a. Grace: Gradient-based active learning with curriculum enhancement for multimodal sentiment analysis. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 5702–5711.

Yafang Li, Zhihua Chu, Caiyan Jia, and Baokai Zu. 2024b. Samgat: structure-aware multilevel graph attention networks for automatic rumor detection. *PeerJ Computer Science*, 10:e2200.

Hongzhan Lin, Jing Ma, Mingfei Cheng, Zhiwei Yang, Liangliang Chen, and Guang Chen. 2021. Rumor detection on twitter with claim-guided hierarchical graph attention networks. *arXiv preprint arXiv:2110.04522*.

Bo Liu, Xiangguo Sun, Qing Meng, Xinyan Yang, Yang Lee, Jiuxin Cao, Junzhou Luo, and Roy Ka-Wei Lee. 2022. Nowhere to hide: Online rumor detection based on retweeting graph neural networks. *IEEE transactions on neural networks and learning systems*, 35(4):4887–4898.

Jing Ma, Wei Gao, Zhongyu Wei, Yueming Lu, and Kam-Fai Wong. 2015. Detect rumors using time series of social context information on microblogging websites. In *Proceedings of the 24th ACM international on conference on information and knowledge management*, pages 1751–1754.

Jing Ma, Wei Gao, and Kam-Fai Wong. 2017. Detect rumors in microblog posts using propagation structure via kernel learning. Association for Computational Linguistics.

Katerina Margatina, Giorgos Vernikos, Loïc Barrault, and Nikolaos Aletras. 2021. Active learning by acquiring contrastive examples. *arXiv preprint arXiv:2109.03764*.

Shengsheng Qian, Jinguang Wang, Jun Hu, Quan Fang, and Changsheng Xu. 2021. Hierarchical multi-modal contextual attention network for fake news detection. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, pages 153–162.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, and 1 others. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PmLR.

Pengzhen Ren, Yun Xiao, Xiaojun Chang, Po-Yao Huang, Zhihui Li, Brij B Gupta, Xiaojiang Chen, and Xin Wang. 2021. A survey of deep active learning. *ACM computing surveys (CSUR)*, 54(9):1–40.

John Schmid Jr. 1947. The relationship between the coefficient of correlation and the angle included between regression lines. *The Journal of Educational Research*, 41(4):311–313.

Ozan Sener and Silvio Savarese. 2017. Active learning for convolutional neural networks: A core-set approach. *arXiv preprint arXiv:1708.00489*.

Lanyu Shang, Yang Zhang, Bozhang Chen, Ruohan Zong, Zhenrui Yue, Huimin Zeng, Na Wei, and Dong Wang. 2024. Mmadapt: A knowledge-guided multi-source multi-class domain adaptive framework for early health misinformation detection. In *Proceedings of the ACM Web Conference 2024*, pages 4653–4663.

Lingzhi Shen, Yunfei Long, Xiaohao Cai, Imran Razzak, Guanming Chen, Kang Liu, and Shoaib Jameel. 2025. Gamed: Knowledge adaptive multi-experts decoupling for multimodal fake news detection. In *Proceedings of the Eighteenth ACM International Conference on Web Search and Data Mining*, pages 586–595.

Meng Shen, Yizheng Huang, Jianxiong Yin, Heqing Zou, Deepu Rajan, and Simon See. 2023. Towards balanced active learning for multimodal classification. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 3434–3445.

Eyal Shnarch, Ariel Gera, Alon Halfon, Lena Dankin, Leshem Choshen, Ranit Aharonov, and Noam Slonim. 2022. Cluster & tune: Boost cold start performance in text classification. *arXiv preprint arXiv:2203.10581*.

Shivangi Singhal, Rajiv Ratn Shah, Tanmoy Chakraborty, Ponnurangam Kumaraguru, and Shin'ichi Satoh. 2019. Spotfake: A multi-modal framework for fake news detection. In *2019 IEEE fifth international conference on multimedia big data (BigMM)*, pages 39–47. IEEE.

Samarth Sinha, Sayna Ebrahimi, and Trevor Darrell. 2019. Variational adversarial active learning. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5972–5981.

Samia Tasnim, Md Mahbub Hossain, and Hoimonty Mazumder. 2020. Impact of rumors and misinformation on covid-19 in social media. *Journal of preventive medicine and public health*, 53(3):171.

Camille Thibault, Jacob-Junqi Tian, Gabrielle Peloquin-Skulski, Taylor Lynn Curtis, James Zhou, Florence Laflamme, Luke Yuxiang Guan, Reihaneh Rabbany, Jean-François Godbout, and Kellin Pelrine. 2025. A guide to misinformation detection data and evaluation. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pages 5801–5809.

Laura Tolosi, Andrey Tagarev, and Georgi Georgiev. 2016. An analysis of event-agnostic features for rumour classification in twitter. In *Proceedings of the*

International AAAI Conference on Web and Social Media, volume 10, pages 151–158.

Toan Tran, Thanh-Toan Do, Ian Reid, and Gustavo Carneiro. 2019. Bayesian generative active deep learning. In *International conference on machine learning*, pages 6295–6304. PMLR.

Constantino Tsallis. 2022. Entropy. *Encyclopedia*, 2(1):264–300.

Yaqing Wang, Fenglong Ma, Zhiwei Jin, Ye Yuan, Guangxu Xun, Kishlay Jha, Lu Su, and Jing Gao. 2018. Eann: Event adversarial neural networks for multi-modal fake news detection. In *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining*, pages 849–857.

Ying Wang, Fuyuan Ma, Zhaoqi Yang, Yaodi Zhu, Bo Yang, Pengfei Shen, and Lei Yun. 2025. Rumor detection with adaptive data augmentation and adversarial training. *Journal of Artificial Intelligence Research*, 82:1175–1204.

Penghui Wei, Nan Xu, and Wenji Mao. 2019. Modeling conversation structure and temporal dynamics for jointly predicting rumor stance and veracity. *arXiv preprint arXiv:1909.08211*.

Yang Wu, Pengwei Zhan, Yunjian Zhang, Liming Wang, and Zhen Xu. 2021. Multimodal fusion with co-attention networks for fake news detection. In *Findings of the association for computational linguistics: ACL-IJCNLP 2021*, pages 2560–2569.

Chengcheng Yu, Jiapeng Zhu, and Xiang Li. 2024. Graphcbal: Class-balanced active learning for graph neural networks via reinforcement learning. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pages 3022–3031.

Feng Yu, Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. 2019. Attention-based convolutional approach for misinformation identification from massive and noisy microblog posts. *computers & security*, 83:106–121.

Fengzhu Zeng, Wenqian Li, Wei Gao, and Yan Pang. 2024. Multimodal misinformation detection by learning from synthetic data with multimodal llms. *arXiv preprint arXiv:2409.19656*.

Wenqiao Zhang, Lei Zhu, James Hallinan, Shengyu Zhang, Andrew Makmur, Qingpeng Cai, and Beng Chin Ooi. 2022a. Boostmis: Boosting medical image semi-supervised learning with adaptive pseudo labeling and informative active annotation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 20666–20676.

Wentao Zhang, Yu Shen, Yang Li, Lei Chen, Zhi Yang, and Bin Cui. 2021a. Alg: Fast and accurate active learning framework for graph convolutional networks. In *Proceedings of the 2021 international conference on management of data*, pages 2366–2374.

Wentao Zhang, Zhi Yang, Yexin Wang, Yu Shen, Yang Li, Liang Wang, and Bin Cui. 2021b. Grain: Improving data efficiency of graph neural networks via diversified influence maximization. *arXiv preprint arXiv:2108.00219*.

Yuheng Zhang, Hanghang Tong, Yinglong Xia, Yan Zhu, Yuejie Chi, and Lei Ying. 2022b. Batch active learning with graph neural networks via multi-agent deep reinforcement learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 9118–9126.

Wanqing Zhao, Yuta Nakashima, Haiyuan Chen, and Noboru Babaguchi. 2023. Enhancing fake news detection in social media via label propagation on cross-modal tweet graph. In *Proceedings of the 31st ACM international conference on multimedia*, pages 2400–2408.

Fedor Zhdanov. 2019. Diverse mini-batch active learning. *arXiv preprint arXiv:1901.05954*.

Xinyi Zhou, Jindi Wu, and Reza Zafarani. 2020. : Similarity-aware multi-modal fake news detection. In *Pacific-Asia Conference on knowledge discovery and data mining*, pages 354–367. Springer.

Arkaitz Zubiaga, Maria Liakata, and Rob Procter. 2016. Learning reporting dynamics during breaking news for rumour detection in social media. *arXiv preprint arXiv:1610.07363*.

Arkaitz Zubiaga, Maria Liakata, Rob Procter, Kalina Bontcheva, and Peter Tolmie. 2015. Crowdsourcing the annotation of rumourous conversations in social media. In *Proceedings of the 24th international conference on World Wide Web*, pages 347–353.

A Appendix

A.1 ALMMIT Hyperparameters and Settings

After preprocessing, the dataset was partitioned into 15% training, 70% validation (serving as the unlabeled pool), and 15% testing, with additional threads forming the active learning pool. Each conversation thread contained text, images, and propagation graphs: text was encoded with GPT-2 (768-dim, 100 tokens), images with ViT-Base-Patch16-224 (224×224, 768-dim), and graph structure with a GNN–GAT encoder (two LSTMs, four attention heads, 128-dim output). Supervisor training used AdamW (4×10^{-5} LR, batch size 1 with accumulation to 8, 10 epochs), while the Student was initialized from the Supervisor and trained with AdamW, cosine scheduling, batch size 4, and 30 acquisition rounds under a 650-query budget. Pseudo-label self-training combined class-balanced CE, KL divergence, and modality-aware losses with annealing. Results were reported using accuracy, macro-F1, and per-class F1, with all experiments fixed at seed 42 for reproducibility.

A.2 Comparative Models

To evaluate the proposed ALMMIT framework, we compare it against three categories of models: few-label approaches, text-supervised baselines, and multimodal methods. **Few-label approaches.** *Cluster & Tune* inter-trains language models on clustering-based pseudo-labels to boost cold-start performance (Shnarch et al., 2022). *IDoFew* has enhanced this idea by employing dual-stage clustering (SIB → KMeans), improving the reliability of pseudo-labeling (Alsuhaibani et al., 2024). *FLAL* integrates few-label learning with margin-sampling active learning, enabling efficient annotation while achieving strong misinformation detection results on Pheme, Twitter15/16, and AraFacts (Alalawi et al., 2025). **Text-supervised approaches.** *ACAMI*(Yu et al., 2019) has taken into consideration content temporal co-attention for event-level representation(Yu et al., 2019), while *HGCN-RNN* introduces a hierarchical multi-task framework capturing stance and conversation structure(Wei et al., 2019). *Bi-GCN* (Bian et al., 2020)adopts bidirectional graph convolutional networks to model propagation and dispersion, and *StA-PLAN* (Khoo et al., 2020)leverages transformer-based attention for long-distance relational modeling. *RDLNP*(Lao et al., 2021) combines linear sequence learning and nonlinear structure learning with attention dependencies, whereas *DynGCN* (Choi et al., 2021)employs temporal snapshots with GAT layers and Bi-LSTMs for dynamic propagation modeling. **Multimodal approaches.** *EANN*(Wang et al., 2018) adversarially learns event-invariant multimodal features, *SpotFake* (Singhal et al., 2019)concatenates pre-trained text and image embeddings, and *MVAE*(Khattar et al., 2019) employs a bimodal variational autoencoder with a binary classifier. *SAFE*(Zhou et al., 2020) tried to learn cross-modal similarity, while *MCAN*(Wu et al., 2021) and *HMCAN*(Qian et al., 2021) exploit co-attention and hierarchical contextual attention to predict inter- and intra-modal relationships. More recent *CLIP*-based detectors leverage large-scale vision language pretraining for misinformation classification (Radford et al., 2021). In addition, FCN-LP detects misinformation by connecting tweets via multimodal similarity (CLIP), propagating labels across them with positive/negative correlations, and regularizing for unseen events(Zhao et al., 2023). We selected these models because they represent competitive baselines across few-label learning, text-supervised propagation modeling, and multimodal detection, which directly align with the challenges that ALMMIT is designed to address.

Table 4: Final test set results for Ablation A1 (removing modality-aware loss). Numbers outside parentheses show the A1 scores; values in parentheses indicate the baseline ALMMIT (with modality-aware loss). Arrows denote whether performance improved (↑) or degraded (↓) compared to baseline.

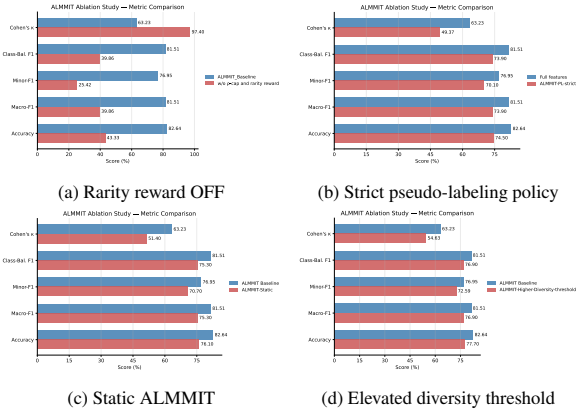
Metric	A1 w/o modality-aware loss vs. ALMMIT
Accuracy	75.98 ↓ (82.64)
Macro-F1	75.40 ↓ (81.51)
Minor-F1	71.64 ↓ (76.95)
Class-Bal. F1	75.40 ↓ (81.51)
Cohen's κ	52.16 ↓ (63.23)

B Ablation Study

For the ablation study, we have fixed the reference configuration and toggled one mechanism and strategy at a time. All variants were evaluated on the Pheme dataset in terms of accuracy, macro-F1, minority-class F1, and budget efficiency.

B.1 Ablation A1: removing modality-aware loss

The results outlined in table 4 present a comparative evaluation of **Ablation A1**, where the modality-



(a) Rarity reward OFF (b) Strict pseudo-labeling policy

(c) Static ALMMIT (d) Elevated diversity threshold

Figure 3: **ALMMIT Ablation Studies.** Each panel shows the effect of disabling or altering a mechanism: (a) removing rarity reward, (b) enforcing strict pseudo-labeling, (c) replacing adaptive scheduling with a static variant, (d) tightening diversity thresholds.

aware loss was removed from ALMMIT, against the ALMMIT model that retained this component. Across all reported metrics, including Accuracy, Macro-F1, Minor-F1, Class-Balanced F1, and Cohen's κ, the ablated variant exhibits consistent declines in performance, as indicated by the downward arrows. Accuracy decreases from 82.64% in ALMMIT with modality-aware loss to 75.98% under Ablation A1, while Macro-F1 and Minor-F1 fall from 81.51% to 75.40% and from 76.95% to 71.64%, respectively. Similarly, Class-Balanced F1 shows a reduction from 81.51% to 75.40%, and Cohen's κ drops substantially from 63.23% to 52.16%. These results underscore the importance of incorporating the β-weighted modality-aware loss as an integral part of ALMMIT, as its removal negatively impacts both overall predictive Accuracy and balanced class representation, thereby weakening the system's stability and reliability.

B.2 Ablation A2: disabled Rarity Reward

In Ablation A2 see Figure 3a, where rarity reward was disabled and only class-proportion gating (ρ-cap) was applied, ALMMIT suffered a severe performance collapse. Macro-F1 dropped to 40% and Accuracy to 43%, with Minority-F1 falling below 26%, indicating that the model largely ignored rare classes. The per-class breakdown confirmed this: recall for one class approached 99% with poor precision, while the other class had recall under 15%. This pattern indicates that the model is biased towards the dominant class. The experiment highlights that while ρ-cap can mitigate overrepresentation, it cannot substitute for rarity-aware reward; instead, it is considered a complement to rarity reward, which explicitly reinforces minority gains. However, this approach can lead to a skewed training signal, resulting in degraded overall performance.

B.3 Ablation A3: Strict pseudo-label consistency

In Figure 3b, we evaluate enforcing strict pseudo-label consistency weighting, where ALMMIT's performance dropped noticeably compared with the full-feature baseline. On the final test set, the model reached a Macro-F1 of 73.9%, Accuracy of 74.5%, and Minor-F1 of 70.1%, all well below the baseline scores (Macro-F1 $\approx$ 81.5%, Accuracy $\approx$ 82.6%, Minor-F1 $\approx$ 77.0%). Looking at the per-class results, Class 0 recall fell sharply to 67.6% (down from 81.4%), which lowered its F1 to 77.8%. By contrast, Class 1 recall increased slightly to 87.8% (from 85.1%), but this came at the expense of precision, which fell to 58.3% (from 70.2%), yielding an F1 of 70.1%. In other words, the model leaned toward higher recall for the minority class but introduced more false positives and lost balance across classes.These results suggest that tightening pseudo-label consistency too much makes the student over-rely on the teacher's confident predictions, while ignoring harder or ambiguous cases that normally provide valuable learning signal.

B.4 Ablation A4: Static query number and confidence threshold

Using static query number and confidence threshold while disabling momentum and round-annealed scheduling removes the feedback loops that normally couple selection pressure to validation dynamics and pool statistics. In the ALMMIT Baseline, the reinforcement learning policy would adaptively adjust (q, τ) via a floor, momentum scaling, and eligibility filters, allowing it to relax and smooth the confidence threshold τ when learning improves or tighten it when pseudo-label noise or class skew increases. By contrast, the ablation with fixed $q = 0.12$ and $\tau = 0.66$ constrained the eligible set regardless of pool dynamics and weakened the diversity guard. As a result, performance was significantly below the ALMMIT equipped with dynamic confidence threshold and momentum scaling. The chart 3c has shown that Accuracy dropped from **82.6%** to **76.1%**, Macro-F1 from **81.5%** to **75.3%**, Minority-F1 from **77.0%** to **70.7%**, and Cohen's κ by nearly twelve points

Table 5: Comparison of the Supervisor baseline and ALMMIT on the Pheme dataset. Higher is better.

Method	Acc	Real News			Fake News		
		P	R	F1	P	R	F1
Supervisor	78.73	82.97	85.25	84.10	69.79	66.08	67.88
ALMMIT	**82.64**	**91.38**	81.35	**86.08**	**70.19**	**85.13**	**76.94**
Δ	+3.91	+8.41	-3.90	+1.98	+0.40	+19.05	+9.06

(63.2% $\rightarrow$ 51.4%). These findings suggest that the dynamic q/τ schedule and momentum-aware gating play a crucial role, delivering better calibration, earlier and higher peaks, and a more balanced error profile than the fixed alternative.

B.5 Ablation A5: Aggressive diversity threshold

The ablation experiment with an elevated diversity threshold highlights both the benefits and drawbacks of enforcing stronger diversity constraints during acquisition. As illustrated on Figure 3d On the final ablation test set, performance decreased relative to the reference configuration: Accuracy fell to **77.7%** (-4.9 points), Macro-F1 to **76.9%** (-4.6 points), Minority-F1 to **72.6%** (-4.4 points), and Cohen's κ to **54.6%** (-8.6 points). The per-class analysis would reveal the underlying the fact that while minority class recall increase to **86.8%**, its precision dropped to **62.4%**. This can be interpreted as an indication of a higher rate of false positives; conversely, Class 0 maintained strong precision (**91.5%**) but with reduced recall (**73.0%**). This precision recall imbalance explains the decline in overall macro-averaged scores. These results suggest that while diversity mechanisms play a crucial role in preventing selection collapse, setting the diversity guard too aggressively reduces representativeness and calibration.

B.6 Active Learning Engine Added Value

To quantify the added value of the active learning engine, we compare the baseline Supervisor with the full ALMMIT framework. As shown in Table 5, active learning improves overall accuracy by 3.91 points and substantially strengthens minority-class detection. In particular, fake-news recall increases from 66.08 to 85.13, yielding a 9.06-point gain in fake-news F1.

B.7 RL Controller Ablation Study

To assess the contribution of the reinforcement learning (RL) controller in ALMMIT, we compare it with two alternative strategies for adjusting acquisition and modality weights during the active learning process. The first strategy employs *static weights*, where acquisition and modality parameters remain fixed throughout training. The second strategy follows a *heuristic schedule*, in which acquisition weights gradually transition from uncertainty-driven sampling in early rounds to diversity-oriented sampling in later rounds. In contrast, the proposed RL controller learns a policy that dynamically adapts acquisition weights (α) and modality weights (β) based on reward feedback obtained during training.

Table 6 presents the results on the Pheme dataset under the few-label setting.

Table 6: Comparison of different control strategies for acquisition and modality weighting on the Pheme dataset.

Controller Type	**Accuracy**	**Macro-F1**
Static weights	79.2	80.1
Heuristic schedule	80.3	81.7
RL controller (ALMMIT)	**82.6**	**84.0**

The results indicate that heuristic scheduling improves performance relative to static weighting, while the RL controller achieves the strongest overall performance. These findings suggest that dynamically adapting acquisition strategies and modality weights during training provides advantages over fixed or manually designed policies. By enabling coordinated adjustment of sampling strategies and multimodal fusion across active learning rounds, the RL controller contributes to more effective learning under limited supervision. These observations highlight the benefit of an adaptive control mechanism for jointly optimizing acquisition strategies and modality weighting in multimodal active learning settings.

B.8 Reward Sensitivity Analysis

To evaluate the robustness of the reinforcement learning reward design, we conduct a sensitivity analysis by modifying key components of the reward and training configuration. Specifically, we examine the impact of removing the rarity reward, altering pseudo-label confidence weighting, and modifying the diversity threshold used in active learning sampling. Table 7 summarizes the results on the Pheme dataset. The results show that removing the rarity reward leads to a substantial degradation in performance, particularly on the minor-

Table 7: Sensitivity analysis of reward components and training configuration.

Configuration	Macro-F1	Minor-F1	CB-F1
ALMMIT (Full)	81.5	76.9	81.5
w/o rarity reward	39.8	25.4	39.8
Strict pseudo-label weighting	73.9	69.1	73.9
High diversity threshold	76.9	72.6	76.9

ity class, confirming its importance in mitigating class imbalance during active learning. Adjusting pseudo-label weighting and diversity thresholds results in moderate performance variations but does not destabilize training. These observations indicate that the RL reward structure plays an important role in balancing exploration, class coverage, and pseudo-label reliability.

B.9 Dataset Statistics

Table 8 summarizes the statistical characteristics of the datasets used in our experiments. The Pheme dataset contains 5,802 conversation threads, comprising 3,830 real news instances and 1,972 fake news instances. These threads are further segmented into 37,809 temporal stages, with an average of 6.52 stages per thread. This temporal segmentation reflects the dynamic evolution of online discussions and enables ALMMIT to capture stage-aware propagation patterns during misinformation detection. In contrast, Twitter15 contains 1,458 threads, including 372 real news and 1,086 fake news instances. Since the dataset does not naturally provide temporal stage information, each thread is treated as a single stage in our experiments. Twitter15 also contains no image modality, allowing us to evaluate the robustness of the modality-aware design when visual information is absent.

Overall, the datasets provide complementary characteristics. Pheme includes multimodal information with temporal propagation structure, while Twitter15 offers a larger proportion of fake news instances but lacks visual inputs. These differences enable us to evaluate ALMMIT under both multimodal and missing-modality scenarios.

Table 8: Dataset statistics. Pheme threads are segmented into multiple stages, while Twitter15 is treated as single-stage.

Statistic	**Pheme**	**Twitter15**
Real News	3830	372
Fake News	1972	1086
Total threads	5802	1458
Total stages	37809	1458
Mean stages per thread	6.52	1.00
Min stages per thread	1	1
Max stages per thread	60	1
Mean posts per thread	17.79	28.23
Total posts	103212	41154
Total images	6698	0
Total edges	198390	80850