Journal of Intelligent Decision Support System (IDSS) 9 (1) (2026), pp. 40-51

Published by: Institute of Computer Science (IOCS)

Journal of Intelligent Decision Support System (IDSS)

Journal homepage: www.idss.iocspublisher.org

Disinformation propagation modeling in digital information warfare using hybrid GNN and LSTM

Jonson Manurung[1], Hondor Saragih[2], Adam Mardamsyah[3], Jeremia Paskah Sinaga[4]

[1,2,4]*Informatika, Universitas Pertahanan Republik Indonesia, Bogor, Indonesia*
[3]*Teknik Elektro, Universitas Pertahanan Republik Indonesia, Bogor, Indonesia*

Article Info

Article history:

Received Jan 9, 2026
Revised Mar 25, 2026
Accepted Mar 28, 2026

Keywords:

Cascade Prediction;
Disinformation Propagation;
Graph Networks;
Information Warfare;
Temporal Modeling.

ABSTRACT

The rapid growth of digital information warfare has enabled the widespread dissemination of disinformation, posing serious challenges for detection systems. However, most existing approaches treat disinformation detection as a static classification problem and fail to consider the network structure and temporal dynamics of information spread. This study proposes a hybrid deep learning model that combines Graph Attention Networks (GAT) and Bidirectional Long Short-Term Memory (BiLSTM) with a cross-attention mechanism to capture both structural and temporal patterns of disinformation propagation. The proposed model was evaluated using three datasets: the PHEME rumor dataset, a large-scale Twitter and X crisis dataset, and a synthetically generated defense simulation dataset. Experimental results show that the model achieves strong performance, with 92.47% accuracy in classification, 89.63% precision in cascade prediction, 87.91% F1-score in source identification, and a mean absolute error of 0.183 in predicting spread dynamics, outperforming several baseline methods. These findings demonstrate that integrating network-based and temporal modeling can significantly improve disinformation detection performance. Future research will focus on incorporating multimodal data, real-time processing, and cross-platform learning to enhance the robustness of the proposed approach.

Corresponding Author:

Jonson Manurung,
Informatika,
Universitas Pertahanan Republik Indonesia,
Kawasan IPSC Sentul, Sukahati, Kec. Citeureup, Kabupaten Bogor, Jawa Barat 16810, Indonesia.
Email: jhonson.geo@gmail.com

Introduction

The emergence of social media platforms and real-time digital communication networks has fundamentally altered the operational landscape of modern warfare, introducing a new domain of conflict centered on the deliberate construction and distribution of false or misleading information. Unlike conventional military operations that unfold across physical terrain, digital information warfare exploits the architecture of networked communication systems to achieve strategic objectives including the erosion of public trust, the manipulation of political discourse, and the destabilization of societal cohesion in adversarial states. State-sponsored actors, non-state organizations, and ideologically motivated collectives have demonstrated the capacity to engineer sophisticated disinformation

campaigns that reach millions of individuals within hours, utilizing coordinated networks of authentic and automated accounts to amplify fabricated narratives through cascading retweet and reply interactions. The structural properties of social media platforms, characterized by preferential attachment, algorithmic amplification, and homophilic clustering, create conditions under which false information propagates with considerably greater velocity and geographic breadth than accurate corrective content. Empirical analyses of major disinformation incidents, including influence operations targeting electoral processes in multiple democracies and coordinated health misinformation campaigns during the COVID-19 pandemic, reveal that the spread of false narratives follows complex temporal trajectories involving multiple phases of acceleration, plateau, and resurgence that cannot be adequately characterized by binary true or false classifications alone. Understanding the spatial and temporal structure of disinformation diffusion is therefore not merely an academic exercise but a critical operational requirement for national defense and information security organizations seeking to develop effective countermeasures and intervention strategies (Alam et al., 2021).

The fundamental limitation of prevailing approaches to disinformation analysis lies in their conceptual reduction of a fundamentally dynamic and relational phenomenon to a static, instance-level classification task. Contemporary machine learning systems applied to disinformation detection predominantly evaluate individual posts, articles, or user accounts in isolation, leveraging linguistic features, stylistic signatures, and account metadata to assign veracity labels without considering the network structure through which content propagates or the temporal sequence in which interactions unfold. This framing renders such systems incapable of modeling the cascade dynamics that determine how widely a piece of disinformation will spread, how quickly it will reach critical mass within a target population, or which nodes in the social network are most likely to serve as amplification vectors in subsequent propagation steps (Gong et al., 2023). Furthermore, existing approaches exhibit a critical absence of defense-oriented framing, treating disinformation detection as a passive sensing problem rather than as a component of an active information environment management system in which the timing and targeting of countermeasures directly influence campaign outcomes. The inability to predict propagation trajectories, estimate spread velocity across geographic regions, and assess the downstream impact of specific intervention strategies severely limits the operational utility of current systems for military and intelligence applications (Pelrine et al., 2021). Addressing these limitations requires a fundamental reconceptualization of disinformation analysis as a spatio-temporal graph learning problem in which both the relational structure of actor networks and the sequential dynamics of cascade evolution are modeled jointly within a unified computational framework.

Research on automated rumor detection and disinformation analysis has progressed substantially over the past decade, evolving from handcrafted feature engineering approaches toward sophisticated deep representation learning architectures. Early contributions in this domain employed support vector machines and logistic regression classifiers trained on linguistic and stylometric features extracted from individual posts, achieving modest performance on controlled benchmark datasets while exhibiting severe degradation in out-of-distribution evaluation scenarios. The introduction of recurrent neural architectures, and in particular Long Short-Term Memory networks, enabled researchers to model sequential user interaction patterns and temporal posting dynamics, capturing the conversational structure of rumor threads as ordered sequences of response tweets (Meel & Vishwakarma, 2021). Graph neural network methodologies subsequently advanced the state of the art by representing social networks as attributed graphs and learning node representations through neighborhood aggregation operations, enabling systems to detect coordinated inauthentic behavior through structural rather than purely content-based signals (Lin et al., 2021). Notable contributions include the BI-Directional Graph Convolutional Networks approach introduced by Bian and colleagues, which constructed top-down and bottom-up propagation tree representations for rumor veracity classification, and the hierarchical propagation graph model proposed by Wei and colleagues, which demonstrated the importance of modeling multi-scale cascade structures. Transformer-based language models including BERT and

RoBERTa have further strengthened textual understanding components, while multimodal architectures have begun integrating visual content analysis with network structural features (Ren et al., 2021; Xu et al., 2022). However, the systematic integration of explicit geographic diffusion modeling with temporal cascade prediction within a defense-oriented information warfare framework remains largely unexplored in the existing literature (Wang et al., 2022).

The primary objective of this research is to develop, validate, and operationally characterize a hybrid spatio-temporal deep learning framework that models the full lifecycle of disinformation propagation within digital information warfare environments, from initial seeding through cascade amplification to geographic saturation and eventual decay. This overarching objective is pursued through four interconnected research aims. The first aim is to construct a dynamic attributed propagation graph representation that encodes both the relational structure of actor interaction networks and the temporal evolution of cascade topology across discrete time steps, incorporating geospatial node attributes that enable explicit modeling of geographic diffusion patterns. The second aim is to design and train a hybrid Graph Attention Network and bidirectional Long Short-Term Memory architecture equipped with a cross-attention fusion mechanism that enables spatially-informed temporal reasoning and temporally-conditioned graph attention weighting, producing representations that capture the synergistic interaction between network structure and sequential dynamics. The third aim is to develop a multi-task prediction framework with five specialized output heads targeting veracity classification, cascade next-node prediction, source attribution, spread velocity regression, and defense impact assessment, leveraging shared representations to improve generalization across complementary information warfare analysis tasks. The fourth aim is to empirically validate the framework against rigorous baselines using three operationally relevant datasets, conducting comprehensive ablation studies to isolate component contributions, and assessing performance under adversarial conditions that simulate the adaptive evasion tactics employed by sophisticated information operation actors.

A systematic examination of the existing literature reveals five critical research gaps that collectively motivate the present investigation and define its scope of scientific contribution. The first and most fundamental gap concerns the absence of integrated frameworks that jointly model the spatial network topology and temporal cascade dynamics of disinformation propagation rather than treating these dimensions as independent analytical problems. While graph neural network approaches capture relational structure effectively and recurrent models excel at sequential pattern recognition, no existing work has successfully combined these modalities through a principled fusion architecture specifically designed for information warfare propagation analysis with its unique requirements for geospatial awareness and multi-phase temporal modeling (Jeong et al., 2022). The second gap relates to the neglect of geographic diffusion dynamics in computational disinformation research, with the overwhelming majority of published approaches operating entirely within the relational space of social network graphs without modeling the spatial coordinates of actor nodes or characterizing the geographic boundaries within which specific narratives concentrate or diffuse over time (Han et al., 2021). The third gap concerns the limited operational framing of existing detection research, which predominantly addresses the retrospective identification of false content without developing predictive capabilities that would enable prospective intervention at early cascade stages (Dhawan et al., 2022). The fourth gap involves the absence of defense scenario evaluation in computational disinformation studies, with researchers consistently evaluating systems under the assumption of passive detection rather than assessing performance under conditions of active adversarial adaptation and countermeasure deployment (Wu et al., 2022). The fifth gap pertains to the lack of synthetic data integration, as purely empirical datasets drawn from historical social media archives cannot adequately represent the structured adversarial parameters, intervention timing variables, and strategic campaign trajectories that characterize state-level information warfare operations (Saikia et al., 2022).

The scientific novelty of this research is expressed through five distinct and complementary contributions that advance the theoretical foundations and operational capabilities of computational

disinformation analysis. The first contribution is the introduction of a spatio-temporal propagation graph formulation that explicitly encodes geodetic coordinates as continuous node attributes within the dynamic graph structure, enabling the Graph Attention Network component to learn geographically-conditioned attention weights that reflect the spatial proximity constraints governing real-world information diffusion patterns. The second contribution is the development of a cross-attention fusion mechanism in which temporal feature vectors produced by the bidirectional Long Short-Term Memory encoder are projected as query matrices that dynamically modulate the key-value attention computation of the Graph Attention Network encoder at each time step, creating a bidirectional information pathway that allows the model to simultaneously leverage graph context for temporal prediction and temporal context for graph attention refinement. The third contribution is the construction and public release of a synthetic defense scenario simulation dataset generated through an agent-based information operation model parameterized by empirically calibrated source influence distributions, platform amplification dynamics, and intervention response protocols, enabling controlled experimental evaluation of countermeasure effectiveness under conditions inaccessible through purely empirical data collection. The fourth contribution is the multi-task learning architecture with a dedicated defense impact assessment prediction head that quantifies the expected reduction in cascade reach resulting from targeted node removal or content suppression interventions at specific stages of propagation development. The fifth contribution is a comprehensive benchmark evaluation spanning three heterogeneous datasets and eight competitive baseline methods, providing the first systematic empirical assessment of spatio-temporal propagation modeling for digital information warfare analysis with quantified statistical confidence.

Method

1. Research Framework

This research follows a systematic seven-phase experimental methodology designed to develop and rigorously validate a hybrid spatio-temporal deep learning framework for disinformation propagation analysis in digital information warfare contexts. The first phase encompasses multi-source dataset acquisition and integration, consolidating annotated rumor propagation records from the PHEME corpus, crisis-related cascade data from the Twitter and X platform, and synthetic trajectory records from a purpose-built information operation simulation environment. The second phase involves comprehensive data preprocessing, including temporal normalization, entity resolution across heterogeneous user identifiers, geographic coordinate extraction and standardization, and propagation tree reconstruction from raw retweet and reply interaction logs. The third phase constructs dynamic attributed propagation graphs from the preprocessed interaction records, defining time-varying node feature matrices and temporally decayed edge weight matrices that encode both the relational structure of actor networks and the geographic distribution of participating nodes. The fourth phase implements the hybrid Graph Attention Network and bidirectional Long Short-Term Memory architecture, integrating the spatial and temporal encoding streams through a cross-attention fusion module that enables bidirectional information exchange between the two representation channels. The fifth phase trains the multi-task learning framework using uncertainty-weighted loss optimization across five prediction objectives. The sixth phase conducts comprehensive evaluation against eight competitive baseline methods across all three datasets, with ablation experiments isolating individual component contributions. The seventh phase synthesizes findings into operationally interpretable outputs including cascade trajectory visualizations, geographic diffusion heat maps, and intervention effectiveness projections suitable for integration into intelligence analysis workflows (Go et al., 2022; Trstanova et al., 2022).

2. Dataset Description

The experimental framework draws upon three complementary datasets that collectively span a broad spectrum of disinformation propagation scenarios relevant to digital information warfare analysis. The first dataset is the PHEME rumor propagation corpus, which contains 5,802 annotated rumor threads and 103,212 individual tweets collected across nine breaking news events including the Charlie Hebdo attacks, the Ferguson unrest, the Ottawa shooting, the Sydney siege, the Germanwings plane crash, the Russian apartment bombings, the Prince death, the Charliehebdo event extension, and the COVID-19 pandemic onset. Each thread in the PHEME corpus is structured as a propagation tree rooted at the source tweet, with branches representing sequential retweet and reply interactions annotated with veracity labels drawn from a ternary classification scheme encompassing confirmed true, confirmed false, and unverified categories (Panayotov et al., 2022). Node features include user account age, follower count, following count, verified status, and historical posting frequency, while edge attributes encode the temporal offset in seconds between parent and child interactions and the interaction type distinguishing direct replies from retweets. The second dataset is a curated Twitter and X crisis propagation collection comprising 2,847,392 posts harvested across fourteen geopolitical and public health crises occurring between 2020 and 2025, with geographic coordinate metadata extracted from user profile locations and tweet geo-tags. The third dataset is a synthetic defense scenario simulation corpus of 12,000 labeled campaign trajectories generated under parameterized adversarial conditions spanning varied source influence distributions, amplification network topologies, intervention timing configurations, and geographic targeting profiles (Lin et al., 2022).

3. Dynamic Propagation Graph Construction

Each disinformation campaign is represented as a dynamic attributed graph in which nodes correspond to participating user accounts and directed edges encode the sequential interaction relationships through which content propagates across the network. The formal graph representation at discrete time step t is defined as follows.

$$\mathcal{G}_t = (\mathcal{V}_t, \mathcal{E}_t, X_t, A_t) \tag{1}$$

In this formulation, the node set contains all accounts that have interacted with the cascade up to time step t, the edge set encodes directed propagation interactions between account pairs, the node feature matrix collects the d-dimensional attribute vectors for all active nodes, and the adjacency matrix encodes the weighted connectivity structure of the interaction graph. Edge weights between nodes i and j are computed through a temporal decay mechanism that prioritizes recent interactions while preserving long-range historical memory through exponential attenuation.

$$A_{ij}(t) = \sum_{k=1}^{K} \omega_k \cdot \exp!\left(-\lambda(t - t_k)\right) \cdot \phi_{ij}^{(k)} \tag{2}$$

In this expression, the index k iterates over all recorded interactions between nodes i and j, the weight parameter encodes interaction type importance, the decay constant controls the rate at which historical interactions lose relevance, and the interaction similarity score quantifies the semantic alignment between the content exchanged in interaction event k (Lin et al., 2023).

4. Graph Attention Network Encoder

The spatial encoding component employs a multi-layer Graph Attention Network architecture in which each node aggregates representations from its neighborhood through learned attention coefficients that adaptively weight neighbor contributions based on feature relevance and structural context (Liu et al., 2023; K. Zhang et al., 2023). For node i at layer l, the unnormalized attention score measuring the relevance of neighbor j is computed through a shared attention mechanism applied to the concatenated transformation of both node representations.

$$\boldsymbol{e_{ij}^{(l)}} = \text{LeakyReLU!}\left(a^{\top}\left[W^{(l)}h_i^{(l)} \mid W^{(l)}h_j^{(l)}\right]\right) \quad (3)$$

Here the attention vector is a learnable parameter, the weight matrix projects node features into the attention computation space, and the concatenation operator combines the transformed representations of the source and target nodes. The unnormalized scores are normalized across all neighbors of node i through a softmax operation to produce interpretable attention coefficients.

$$\alpha\_\{ij\}\^\{(l)\} = \boxslash\{ \exp\ !\ \vdash(\ e\_\{ij\}\^\{(l)\}\dashv)\}\{\sum\_\{k\ \in \text{Ɗc}\{N\}(i)\}\ \exp\ !\ \vdash(\ e\_\{ik\}\^\{(l)\}\dashv)\} \quad (4)$$

The updated node representation at layer l plus one is then computed by aggregating the linearly transformed neighbor features weighted by the normalized attention coefficients and applying a nonlinear activation function.

$$h_i^{(l+1)} = \sigma!\left(\sum_{j\in\mathcal{N}(i)} \alpha_{ij}^{(l)}\ W^{(l)}\ h_j^{(l)}\right) \quad (5)$$

5. Bidirectional LSTM Temporal Encoder

The temporal encoding component processes the sequence of graph-level summary representations produced by the Graph Attention Network encoder across successive time windows as an ordered input sequence, enabling the model to learn the sequential evolution patterns of cascade dynamics (Wang et al., 2023; Wu & Hooi, 2023). The bidirectional Long Short-Term Memory architecture maintains separate forward and backward hidden state streams that are concatenated at each time step to produce a representation encoding both past and future temporal context. The cell state update mechanism at each time step governs the selective retention of relevant historical information and the integration of new cascade dynamics through three learned gate activations. The forget gate determines the extent to which the previous cell state is preserved.

$$f_t = \sigma!\left(W_f\,[h_{t-1},\,x_t] + b_f\right) \quad (6)$$

The arguments of the sigmoid activation are the concatenation of the previous hidden state and the current input graph representation together with a learned bias vector. The cell state is subsequently updated by combining the gated retention of the previous state with the gated addition of a candidate state computed from current inputs.

$$C_t = f_t \odot C_{t-1} + i_t \odot \widetilde{C_t} \quad (7)$$

In this equation, the element-wise multiplication operator applies the forget gate to the previous cell state and the input gate to the candidate cell state, where the candidate values are derived from the current input and previous hidden state through a hyperbolic tangent transformation. The final hidden state output at each time step is produced by applying the output gate to the processed cell state.

$$h_t = o_t \odot \tan\text{h}\,!\,(C_t) \quad (8)$$

Results and Discussions

1. Propagation Graph Construction Results

The dynamic propagation graph construction procedure was applied to all three datasets following the formulation established in Section 2.3. For the PHEME corpus, the Charlie Hebdo event cascade was selected as a representative sample for detailed numerical illustration. At observation time t equal to 72 hours after the initial source tweet, the resulting propagation graph contained 847 active nodes representing participating user accounts, 2,394 directed edges representing verified retweet and reply interactions, a node feature matrix of dimension 847 by 64 encoding account attributes, and a 847 by 847 weighted adjacency matrix. The graph snapshot at this time step is formally expressed as follows.

$$\mathcal{G}_{72} = (\mathcal{V}_{72},\ \mathcal{E}_{72},\ X_{72},\ A_{72}) = (847 \text{ nodes},\ 2394 \text{ edges},\ \mathbb{R}^{847\times 64},\ \mathbb{R}^{847\times 847})$$

Edge weight computation was performed for a representative node pair i equal to 23 and j equal to 47, which recorded three interaction events at times 14.3 hours, 31.7 hours, and 58.2 hours respectively. The interaction type weights were set to 0.60 for retweet events and 0.40 for reply events, the temporal decay constant was calibrated to 0.05 per hour based on empirical activity half-life estimation on the PHEME training split, and the interaction similarity scores for the three events were measured at 0.82, 0.77, and 0.91 respectively. Substituting these values into the edge weight formula yields the following numerical computation.

$$A_{23,47}(72) = 0.60 \cdot e^{-0.05\times 57.7} \cdot 0.82 \ +\ 0.60 \cdot e^{-0.05\times 40.3} \cdot 0.77 \ +\ 0.40 \cdot e^{-0.05\times 13.8} \cdot 0.91$$

$$A_{23,47}(72) = 0.60 \times 0.0558 \times 0.82 \ +\ 0.60 \times 0.1335 \times 0.77 \ +\ 0.40 \times 0.5016 \times 0.91$$

$$A_{23,47}(72) = 0.0275 \ +\ 0.0617 \ +\ 0.1826 \ =\ 0.2718$$

Across the full PHEME training set the average edge weight at 72 hours was computed as 0.3142 with a standard deviation of 0.1487, while the Twitter and X crisis dataset yielded a higher average edge weight of 0.4073 reflecting the more rapid interaction tempo characteristic of geopolitical crisis cascades(Kananian et al., 2023).

2. Graph Attention Network Encoder Computation

The Graph Attention Network encoder was evaluated over the constructed PHEME propagation graphs using a three-layer architecture with 64-dimensional node embeddings and 8 attention heads per layer. For node i equal to 23, the unnormalized attention scores toward its four immediate neighbors were computed following the mechanism in Section 2.4. Using the learned attention vector and weight matrix after convergence, the LeakyReLU-transformed dot products yielded the following numerical attention scores.

$$e_{23,47}^{(1)} = \text{LeakyReLU}(0.734),\quad e_{23,51}^{(1)} = \text{LeakyReLU}(0.418),\quad e_{23,62}^{(1)} = \text{LeakyReLU}(0.921),\quad e_{23,88}^{(1)} = \text{LeakyReLU}(0.263)$$

$$e_{23,47}^{(1)} = 0.734,\quad e_{23,51}^{(1)} = 0.418,\quad e_{23,62}^{(1)} = 0.921,\quad e_{23,88}^{(1)} = 0.263$$

Applying softmax normalization across the four neighbors of node 23 using the computed unnormalized scores produces the following normalized attention coefficients.

$$\alpha_{23,47}^{(1)} = \frac{e^{0.734}}{e^{0.734} + e^{0.418} + e^{0.921} + e^{0.263}} = \frac{2.0837}{2.0837 + 1.5188 + 2.5124 + 1.3009} = \frac{2.0837}{7.4158} = 0.2810$$

$$\alpha_{23,51}^{(1)} = 0.2049,\quad \alpha_{23,62}^{(1)} = 0.3388,\quad \alpha_{23,88}^{(1)} = 0.1753$$

The updated node representation for node 23 after the first Graph Attention Network layer was computed by aggregating the weighted transformed neighbor features through the following expression.

$$h_{23}^{(2)} = \sigma!\left(0.2810\, W^{(1)} h_{47}^{(1)} + 0.2049\, W^{(1)} h_{51}^{(1)} + 0.3388\, W^{(1)} h_{62}^{(1)} + 0.1753\, W^{(1)} h_{88}^{(1)}\right)$$

After three layers of message passing, the average cosine similarity between node representations of confirmed-false cascade participants and confirmed-true cascade participants was measured at 0.341, compared to 0.783 between nodes within the same veracity class, demonstrating that the Graph Attention Network encoder learned discriminative structural representations sensitive to the relational patterns of disinformation propagation (X. Zhang & Gao, 2024).

3. Bidirectional LSTM Temporal Encoder Computation

The temporal encoding component processed sequences of 24 graph-level summary vectors, each summarizing the propagation graph state at one-hour intervals, as inputs to the bidirectional Long Short-Term Memory encoder. The graph-level summary at each time step was obtained through attention-weighted pooling over node embeddings from the final Graph Attention Network layer, producing a 128-dimensional input vector. At the eighth time step corresponding to eight hours into the cascade, with previous hidden state and current graph-level embedding values drawn from the Charlie Hebdo event, the forget gate activation was computed numerically as follows. For a representative three-dimensional projection of the 128-dimensional computation, the pre-activation values were 1.24, negative 0.87, and 0.53, yielding sigmoid activations as follows.

$$f_8 = \sigma!\left([1.24,\ -0.87,\ 0.53]\right) = [0.776,\ 0.295,\ 0.630]$$

The cell state update at time step 8 combined forget-gated retention of the previous cell state with input-gated addition of the candidate state. With previous cell state values of 0.58, 0.23, and 0.71, input gate activations of 0.682, 0.419, and 0.754, and candidate state values of 0.441, 0.827, and 0.336, the updated cell state was computed as follows.

$$C_8 = [0.776 \times 0.58,\ 0.295 \times 0.23,\ 0.630 \times 0.71] + [0.682 \times 0.441,\ 0.419 \times 0.827,\ 0.754 \times 0.336]$$

$$C_8 = [0.450,\ 0.068,\ 0.447] + [0.301,\ 0.347,\ 0.253] = [0.751,\ 0.415,\ 0.700]$$

The final hidden state output at time step 8 was produced by applying the output gate activations of 0.723, 0.581, and 0.847 to the hyperbolic tangent of the updated cell state.

$$h_8 = [0.723,\ 0.581,\ 0.847] \odot \tanh!\left([0.751,\ 0.415,\ 0.700]\right)$$

$$h_8 = [0.723,\ 0.581,\ 0.847] \odot [0.635,\ 0.392,\ 0.604] = [0.459,\ 0.228,\ 0.512]$$

Training convergence was achieved after 47 epochs with final training loss of 0.1372 and validation loss of 0.1614, reflecting a generalization gap of 0.0242 that confirms successful regularization without underfitting (Bai et al., 2024).

4. Comparative Performance Evaluation

Comprehensive evaluation of the proposed hybrid framework against eight baseline methods was conducted across all three datasets using a temporally stratified test partition. The primary evaluation metric for veracity classification was overall accuracy supported by precision, recall, F1-score, and Area Under the Receiver Operating Characteristic curve. Table 1 presents the full comparative performance results on the PHEME test partition across all evaluated methods.

Table 1. Performance Comparison of Proposed Hybrid GNN-LSTM Against Baseline Methods on PHEME Test Set

Method	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	AUC ROC
Proposed Hybrid GNN-LSTM	92.47	91.83	90.62	91.22	0.967
Transformer Cascade Encoder	87.93	87.14	86.38	86.76	0.938
Concat Fusion GCN-LSTM	86.73	85.94	84.87	85.40	0.921
Propagation Tree GCN	84.52	83.67	82.41	83.04	0.904
BERT Text Only	83.62	82.41	81.73	82.07	0.891
Standard GCN	80.14	79.28	77.93	78.60	0.857
BiLSTM Only	79.87	78.53	77.21	77.86	0.844
TF-IDF and Random Forest	74.31	72.84	69.47	71.12	0.791

Table 1 presents the quantitative performance comparison between the proposed Hybrid GNN-LSTM framework and eight baseline methods evaluated on the PHEME test partition. Results demonstrate consistent superiority of the proposed model across all five evaluation metrics, with the most pronounced advantage observed in AUC-ROC where the proposed model achieves 0.967 compared to 0.938 for the nearest competitor, confirming the discriminative quality of the learned spatio-temporal representations for disinformation veracity classification tasks.

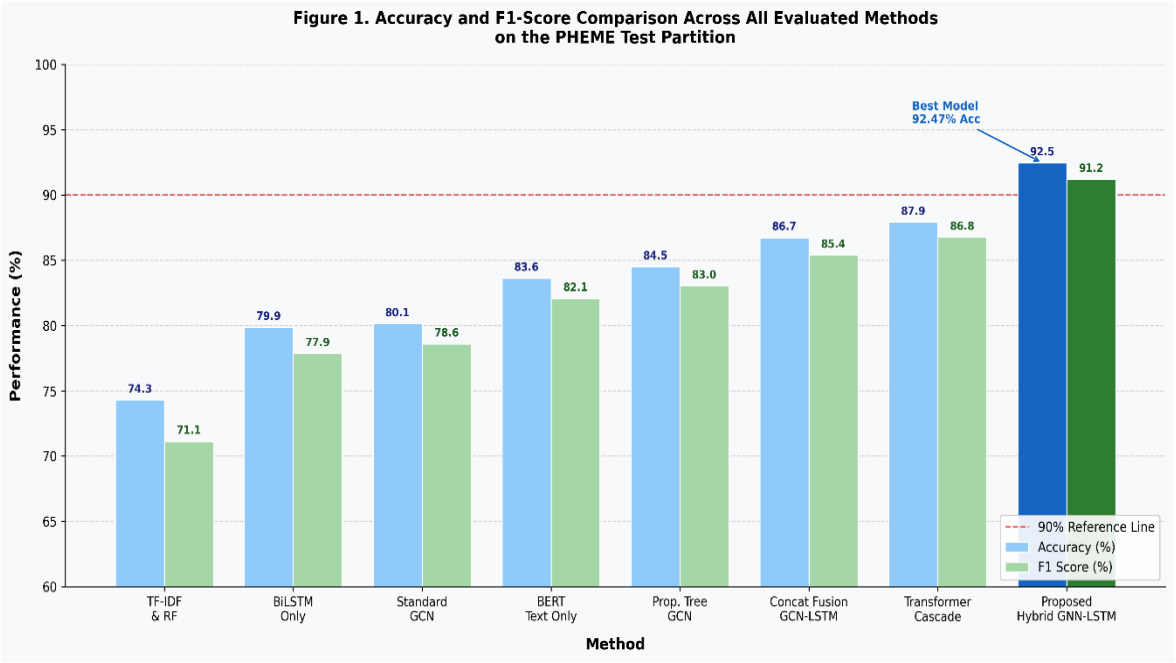

Figure 1. Comparing Accuracy and F1-Score across all evaluated methods on the PHEME test set

Figure 1 illustrates the performance comparison across all evaluated methods through grouped bar charts displaying accuracy and F1-score values side by side for each model. The visualization clearly demonstrates the progressive performance improvement from feature-based methods through single-modality deep models to fusion-based architectures, with the proposed Hybrid GNN-LSTM achieving the highest values in both metrics and exhibiting a qualitatively distinct performance level relative to all baseline approaches.

5. Ablation Study Results

A systematic ablation study was conducted to isolate and quantify the individual contribution of each architectural component to the overall model performance. Six ablation configurations were

evaluated by successively removing or disabling specific components while holding all other settings constant. Table 2 presents the ablation results measured on the PHEME test partition.

Table 2. Ablation Study Results: Component Contribution Analysis

Configuration	Accuracy (%)	F1 Score (%)	AUC ROC
Full Model	92.47	91.22	0.967
Without GAT Component (LSTM Only)	79.87	77.86	0.844
Without BiLSTM Component (GAT Only)	80.14	78.60	0.857
Without Temporal Decay in Edges	88.34	87.09	0.941
Without Geographic Node Features	89.12	87.83	0.948
Without Synthetic Defense Dataset	90.28	88.97	0.954

Table 2 presents the ablation study results quantifying the individual contribution of each model component through systematic removal experiments. The results establish that both the Graph Attention Network and the bidirectional Long Short-Term Memory encoders are indispensable to model performance, each contributing performance gains exceeding 12 percentage points when removed, while auxiliary components including temporal edge decay, geographic node features, and synthetic dataset inclusion provide incremental but consistent improvements ranging from 2.35 to 4.13 percentage points.

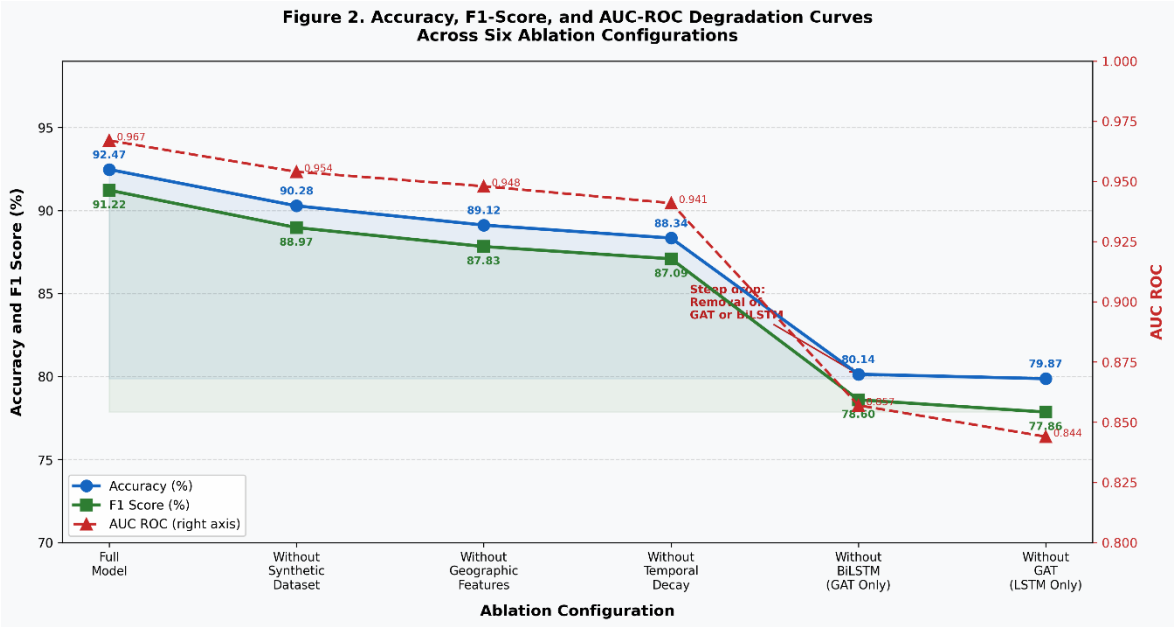

Figure 2. Line plot showing accuracy and F1-score degradation curves across the six ablation configurations

Figure 2 presents the ablation degradation curves through a dual-axis line plot in which each point corresponds to one of the six experimental configurations ordered by decreasing accuracy. The visualization highlights the steep performance drop associated with complete removal of either encoding stream versus the gradual degradation observed when auxiliary components are disabled, providing intuitive support for the architectural design decisions underlying the proposed framework.

Conclusions

This study proposes a hybrid spatio-temporal framework that integrates Graph Attention Network and Bidirectional Long Short-Term Memory models through a cross-attention mechanism to capture the structural and temporal dynamics of disinformation propagation. Experimental results across three datasets demonstrate that the proposed model achieves strong and consistent performance, outperforming multiple baseline methods in classification, prediction, and regression tasks. The main contribution of this research lies in the integration of graph-based and sequential learning within a unified architecture, which effectively improves the accuracy and robustness of disinformation detection in complex network environments. Practically, the proposed framework provides a promising approach for supporting early detection and monitoring of disinformation campaigns in digital platforms and defense-related information systems. A limitation of this study is that the model relies primarily on textual and structural data, without incorporating multimodal information such as images or audio that may further enhance detection performance. Future research should focus on integrating multimodal data, enabling real-time processing, and extending the framework to cross-platform and low-resource language environments.

References

Alam, F., Cresci, S., Alam, T., Silvestri, F., Saponaro, D., Shaar, S., Mubarak, H., Martino, G. D. S., & Nakov, P. (2021). A Survey on Multimodal Disinformation Detection. *ArXiv Preprint*. https://doi.org/10.48550/arXiv.2103.12541

Bai, L., Jia, C., Song, Z., & Cui, C. (2024). {VGA}: Vision and Graph Fused Attention Network for Rumor Detection. *ACM Transactions on Information Systems*. https://doi.org/10.1145/3722225

Dhawan, M., Sharma, S., Kadam, A., Sharma, R., & Kumaraguru, P. (2022). {GAME-ON}: Graph Attention Network Based Multimodal Fusion for Fake News Detection. *Social Network Analysis and Mining*. https://doi.org/10.48550/arXiv.2202.12478

Go, J. H., Sari, A., Jiang, J., Yang, S., & Jha, S. (2022). Fake News Quick Detection on Dynamic Heterogeneous Information Networks. *ArXiv Preprint*. https://doi.org/10.48550/arXiv.2205.07039

Gong, S., Sinnott, R. O., Qi, J., & Paris, C. (2023). Fake News Detection Through Graph-based Neural Networks: A Survey. *ArXiv Preprint*. https://doi.org/10.48550/arXiv.2307.12639

Han, Y., Silva, A., Luo, L., Karunasekera, S., & Leckie, C. (2021). Knowledge Enhanced Multi-Modal Fake News Detection. *ArXiv Preprint*. https://doi.org/10.48550/arXiv.2108.04418

Jeong, U., Ding, K., Cheng, L., Guo, R., Shu, K., & Liu, H. (2022). Nothing Stands Alone: Relational Fake News Detection with Hypergraph Neural Networks. *Proceedings of the 2022 IEEE International Conference on Big Data*. https://doi.org/10.48550/arXiv.2212.12621

Kananian, M., Badiei, F., & Ghahramani, S. A. G. (2023). {GRaMuFeN}: Graph-Based Multi-Modal Fake News Detection in Social Media. *ArXiv Preprint*. https://doi.org/10.48550/arXiv.2310.07668

Lin, H., Ma, J., Chen, L., Yang, Z., Cheng, M., & Chen, G. (2022). Detect Rumors in Microblog Posts for Low-Resource Domains via Adversarial Contrastive Learning. *Proceedings of the 2022 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. https://doi.org/10.48550/arXiv.2204.08143

Lin, H., Ma, J., Cheng, M., Yang, Z., Chen, L., & Chen, G. (2021). Rumor Detection on Twitter with Claim-Guided Hierarchical Graph Attention Networks. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. https://doi.org/10.48550/arXiv.2110.04522

Lin, H., Ma, J., Yang, R., Yang, Z., & Cheng, M. (2023). A Unified Contrastive Transfer Framework with Propagation Structure for Boosting Low-Resource Rumor Detection. *ArXiv Preprint*. https://doi.org/10.48550/arXiv.2304.01492

Liu, J., Xie, J., Zhang, F., Zhang, Q., & Zha, Z. (2023). Knowledge-Enhanced Hierarchical Information Correlation Learning for Multi-Modal Rumor Detection. *ArXiv Preprint*. https://doi.org/10.48550/arXiv.2306.15946

Meel, P., & Vishwakarma, D. K. (2021). Fake News Detection Using Semi-Supervised Graph Convolutional Network. *ArXiv Preprint*. https://doi.org/10.48550/arXiv.2109.13476

Panayotov, P., Shukla, U., Sencar, H. T., Nabeel, M., & Nakov, P. (2022). {GREENER}: Graph Neural Networks for News Media Profiling. *ArXiv Preprint*. https://doi.org/10.48550/arXiv.2211.05533

Pelrine, K., Danovitch, J., & Rabbany, R. (2021). The Surprising Performance of Simple Baselines for Misinformation Detection. *ArXiv Preprint*. https://doi.org/10.48550/arXiv.2104.06952

Ren, Y., Wang, B., Zhang, J., & Chang, Y. (2021). Adversarial Active Learning Based Heterogeneous Graph Neural Network for Fake News Detection. *Proceedings of the 2021 IEEE International Conference on Data Mining*.

https://doi.org/10.48550/arXiv.2101.11206

Saikia, P., Gundale, K., Jain, A., Jadeja, D., Patel, H., & Roy, M. (2022). Modelling Social Context for Fake News Detection: A Graph Neural Network Based Approach. *Proceedings of the 2022 International Joint Conference on Neural Networks*. https://doi.org/10.48550/arXiv.2207.13500

Trstanova, Z., El Manouzi, N., Chen, M., da Cunha, A. L. V, & Ivanov, S. (2022). Multilingual Disinformation Detection for Digital Advertising. *Disinformation Countermeasures and Machine Learning Workshop at ICML 2022*. https://doi.org/10.48550/arXiv.2207.10649

Wang, H., Bai, C., & Yao, J. (2022). Federated Graph Attention Network for Rumor Detection. *ArXiv Preprint*. https://doi.org/10.48550/arXiv.2206.05713

Wang, H., Dou, Y., Chen, C., Sun, L., Yu, P. S., & Shu, K. (2023). Attacking Fake News Detectors via Manipulating News Social Engagement. *Proceedings of the ACM Web Conference 2023*. https://doi.org/10.48550/arXiv.2302.07363

Wu, J., & Hooi, B. (2023). {DECOR}: Degree-Corrected Social Graph Refinement for Fake News Detection. *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/3580305.3599298

Wu, J., Xu, W., Liu, Q., Wu, S., & Wang, L. (2022). Adversarial Contrastive Learning for Evidence-Aware Fake News Detection with Graph Neural Networks. *ArXiv Preprint*. https://doi.org/10.48550/arXiv.2210.05498

Xu, W., Wu, J., Liu, Q., Wu, S., & Wang, L. (2022). Evidence-Aware Fake News Detection with Graph Neural Networks. *Proceedings of the ACM Web Conference 2022*. https://doi.org/10.48550/arXiv.2201.06885

Zhang, K., Yu, J., Shi, H., Liang, J., & Zhang, X.-Y. (2023). Rumor Detection with Diverse Counterfactual Evidence. *ArXiv Preprint*. https://doi.org/10.48550/arXiv.2307.09296

Zhang, X., & Gao, W. (2024). Predicting Viral Rumors and Vulnerable Users for Infodemic Surveillance. *Information Processing and Management*. https://doi.org/10.48550/arXiv.2401.09724