



Spatio-Temporal Disinformation Detection Using GNN-LSTM and Cross-Attention for South China Sea Information Analysis

Adam Mardamsyah¹, Marsetio¹, Bambang Irwanto¹, Robby Mochammad Taufik¹

¹ Strategi Pertahanan, Doktoral Ilmu Pertahanan, Universitas Pertahanan Republik Indonesia, Jakarta, Indonesia

Article Info

Article history:

Received Apr 06, 2026

Revised Sep 19, 2026

Accepted Oct 09, 2026

Keywords:

Cross-Attention;
Disinformation Detection;
Long Short-Term Memory;
Spatio-Temporal Graph Neural
Network;
South China Sea.

ABSTRACT

The rapid propagation of disinformation through social media presents a significant challenge because information diffusion is characterized by complex interactions among users and continuously evolving temporal patterns. Conventional deep-learning and transformer-based approaches primarily focus on textual or sequential representations and may not fully capture the structural and temporal characteristics of information propagation. This study proposes a spatio-temporal framework for disinformation detection that integrates Graph Neural Network (GNN)-based spatial representation, Long Short-Term Memory (LSTM)-based temporal modeling, and an adaptive cross-attention mechanism for spatial-temporal fusion. The GNN is employed to model relational dependencies within social-media interaction networks, while LSTM captures the temporal evolution of graph-based representations. Cross-attention is subsequently introduced to adaptively model the interaction between spatial and temporal representations rather than relying solely on fixed feature concatenation. The proposed framework is evaluated using a social-media dataset associated with the South China Sea and compared with CNN, LSTM, GNN, BERT, and GAT-BERT baselines. Experimental results show that the proposed model achieves 95.84% accuracy, 95.30% precision, 95.60% recall, and 95.45% F1-score, outperforming the strongest baseline, GAT-BERT, by 2.63 percentage points in accuracy and 2.55 percentage points in F1-score. The ablation analysis further demonstrates that integrating LSTM with GNN improves accuracy by 5.32 percentage points, while the subsequent incorporation of cross-attention provides an additional improvement of 3.74 percentage points. These findings indicate that adaptive spatial-temporal fusion contributes beyond the independent use of graph and sequential encoders and provides a potential decision-support approach for analyzing dynamic disinformation propagation in geopolitical social-media environments.

This is an open access article under the CC BY-NC license.



Corresponding Author:

Adam Mardamsyah,
Doktoral Ilmu Pertahanan, Strategi Pertahanan,
Universitas Pertahanan Republik Indonesia,
Jl. Salemba Raya No. 14, Jakarta Pusat, 10440, Indonesia
Email: adam.mardamsyah@idu.ac.id

1. INTRODUCTION

The evolution of modern warfare has extended beyond conventional physical domains into increasingly complex digital and cognitive environments, where information can be employed as a

strategic instrument to influence public perception, shape political narratives, and affect decision-making processes [1][2]. Within this environment, disinformation has become an important component of information warfare because digitally mediated information can be produced, amplified, and redistributed at a scale and speed that are difficult to manage through conventional monitoring mechanisms [3]. The increasing dependence on social media has further intensified this problem because information dissemination is no longer determined solely by the content of a message, but also by the relationships among users, interaction patterns, propagation structures, and the temporal evolution of those interactions. Consequently, automated disinformation detection requires analytical approaches capable of representing both the structural characteristics of information networks and their evolution over time [1], [4]–[6].

The South China Sea represents a particularly relevant context for studying information warfare because the region involves overlapping territorial claims, strategic maritime interests, and interactions among multiple state and non-state actors. Information operations in such an environment may involve the coordinated dissemination of narratives across multiple users and communication channels, making the resulting information ecosystem highly dynamic and interconnected [7]–[9]. In contrast to conventional information classification, where individual textual content can be analyzed independently, disinformation propagation in a geopolitical environment is better understood as a networked process. A misleading narrative may originate from one account, be amplified through connected users, and subsequently evolve through replies, mentions, reposts, and other interactions. Therefore, the identification of disinformation should consider not only what information is communicated, but also how information propagates through a network and how that propagation changes over time.

Early approaches to misinformation and disinformation detection predominantly relied on content-based and conventional machine-learning techniques, including linguistic characteristics, sentiment, user characteristics, and textual representations [10][11]. These approaches can be effective when the discriminative information is primarily contained within the content itself. However, they may be insufficient when disinformation is characterized by coordinated propagation behavior or complex interaction structures. In social-media environments, two messages with similar textual characteristics may exhibit substantially different propagation patterns because they are disseminated through different network structures and at different temporal stages. This observation has motivated the development of graph-based approaches that explicitly model relationships among users, messages, and propagation events [12]–[14].

Graph Neural Networks (GNNs) provide a natural representation-learning framework for such relational data because they aggregate information from interconnected nodes and learn representations that reflect local and higher-order structural dependencies [15][16]. In misinformation detection, graph-based approaches can represent users or information entities as nodes and interactions such as reposts, replies, and mentions as edges [17][18]. Previous studies have demonstrated that GNN-based models can improve misinformation detection by exploiting propagation structures that are difficult to capture using purely sequential or content-based models [19]–[21]. However, a conventional GNN is generally designed for a graph representation that is treated as static during a particular learning operation. Such a representation may not adequately capture information diffusion processes in which new users, interactions, and propagation paths continuously emerge.

This limitation has motivated the development of Dynamic Graph Neural Networks (DGNNs) and Temporal Graph Neural Networks (TGNNs), which explicitly incorporate temporal evolution into graph representation learning. Instead of treating a network as a single static graph, dynamic graph methods model a sequence of graph states or temporally ordered interaction events. Several approaches illustrate different strategies for addressing this problem. DySAT (Dynamic Self-Attention Network) jointly applies structural and temporal self-attention to learn node representations from evolving graph structures, allowing structural neighborhood information and temporal dependencies to be modeled within a unified dynamic graph representation [22]. EvolveGCN takes a different

approach by evolving the parameters of the graph convolutional network through time, rather than relying solely on recurrent updates of node embeddings, thereby allowing the GCN itself to adapt as the graph changes [23]. Temporal Graph Networks (TGN) further model dynamic graphs as sequences of time-stamped events and introduce memory mechanisms that allow node representations to evolve as new interactions occur [24]. These developments demonstrate that temporal information is not merely an auxiliary feature but an important dimension of graph representation learning for evolving information networks.

The relevance of dynamic graph learning has also been demonstrated specifically in fake-news and misinformation detection. Temporal propagation-based approaches have been developed to address the limitation of static propagation networks by modeling the continuous evolution of information diffusion. For example, the Temporally Evolving Graph Neural Network for Fake News Detection (TGNF) explicitly models news propagation as a continuous-time dynamic graph and integrates structural, semantic, and temporal information to capture the evolution of news diffusion [32]. These developments indicate that misinformation detection is increasingly moving from static graph classification toward spatio-temporal modeling, in which the topology of an interaction network and the temporal sequence of propagation events are jointly considered.

Another important development is the use of attention-based sequence models. The Transformer architecture demonstrated that attention mechanisms can model dependencies across sequence elements without relying exclusively on recurrent computation [33]. In dynamic graph learning, self-attention has subsequently been employed to capture temporal dependencies, as illustrated by DySAT. Such approaches are particularly attractive because attention can assign different importance to different temporal or structural components. Nevertheless, attention-based temporal modeling and graph-based spatial modeling do not automatically guarantee an effective interaction between the two representations. A model may independently learn structural and temporal representations and subsequently combine them through concatenation or another fixed fusion operation. Such fusion does not necessarily allow the model to determine which structural information is most relevant to a particular temporal state.

This issue is particularly important in disinformation propagation because the relevance of network structure can vary across different stages of information diffusion. A structural pattern that is informative during the initial dissemination of a narrative may not have the same relevance during later amplification or response stages. Similarly, a temporal pattern may be meaningful only when considered in the context of specific network relationships. Therefore, a spatio-temporal architecture should not merely preserve spatial and temporal representations independently; it should enable adaptive interaction between these representations.

Recent research on graph-based misinformation detection has increasingly explored the integration of graph learning with recurrent, attention-based, and transformer architectures [17], [23]–[28]. However, the existing literature indicates several remaining challenges. First, dynamic graph approaches such as DySAT, EvolveGCN, and TGN provide sophisticated mechanisms for modeling temporal graph evolution, but they represent alternative architectural paradigms rather than directly addressing the specific problem of adaptive fusion between independently learned spatial and temporal representations. Second, hybrid graph-sequential approaches may combine GNN and recurrent representations using simple concatenation or independent processing, which does not explicitly model the relevance of one representation with respect to the other. Third, although attention mechanisms have been increasingly adopted in graph and temporal learning, the use of cross-attention specifically as an adaptive mechanism for integrating graph-based spatial representations with sequential temporal representations in disinformation propagation remains insufficiently explored, particularly in the context of geopolitically sensitive information environments.

Based on this research gap, this study proposes a spatio-temporal GNN–LSTM framework with cross-attention fusion for disinformation detection in the South China Sea information environment. The GNN component is used to learn spatial representations from the interaction structure, while the

LSTM component captures temporal dependencies in the evolution of information propagation [25]–[27]. Rather than directly concatenating the two representations, the proposed cross-attention mechanism enables one representation to selectively attend to the informative components of the other representation [27]–[29]. In this formulation, the spatial and temporal representations are therefore not treated as independent features but as interacting representations whose relevance is dynamically determined through attention weights.

The methodological novelty of this study is consequently positioned not in the individual use of GNN or LSTM, since both architectures have been extensively investigated, but in the cross-attention-based interaction mechanism used to adaptively fuse spatial propagation representations and temporal sequence representations. This distinction is important because the objective is not simply to increase architectural complexity by combining two established models, but to investigate whether adaptive spatial-temporal interaction can provide a more informative representation of evolving disinformation propagation [30][31]. The proposed framework is further evaluated against CNN, LSTM, GNN, BERT, and GAT-BERT baselines, together with ablation experiments designed to assess the contribution of the spatial encoder, temporal encoder, and cross-attention fusion mechanism.

The main contributions of this study are as follows. First, a spatio-temporal representation framework is developed for disinformation detection by modeling social-media interactions through graph-based spatial representations and temporal sequential representations. Second, a cross-attention fusion mechanism is introduced to adaptively model the interaction between GNN-derived spatial representations and LSTM-derived temporal representations, rather than relying solely on fixed feature concatenation. Third, a systematic ablation perspective is employed to examine the contribution of individual architectural components and determine whether the observed performance improvement is associated with the proposed spatial-temporal fusion mechanism. Fourth, a comparative experimental evaluation is conducted against conventional deep-learning, graph-based, and transformer-based baseline models to assess the effectiveness of the proposed framework. Finally, a defense-oriented application context is provided by applying the proposed framework to disinformation analysis associated with the South China Sea, where dynamic and coordinated information propagation may constitute a component of broader information-warfare activities.

2. RESEARCH METHOD

2.1. Problem Formulation

Disinformation detection is formulated as a spatio-temporal classification problem over an evolving social-media interaction network. Let the social-media network at time step t be represented as a directed graph $G_t = (V_t, E_t, X_t)$, where V_t denotes the set of users or information-related entities, E_t denotes the set of directed interactions among entities, and $X_t \in \mathbb{R}^{|V_t| \times d_x}$ represents the node-feature matrix at time t . The edge set may represent interactions such as retweets, replies, and mentions, according to the interaction types available in the dataset. The complete temporal network is represented as a sequence of graph snapshots,

$$\mathcal{G} = \{G_1, G_2, \dots, G_T\}, \quad (1)$$

where T denotes the number of temporal observations. For each time step, the spatial structure is encoded using a graph convolutional network, while the resulting sequence of graph representations is processed by an LSTM to capture temporal dependencies. The spatial and temporal representations are subsequently integrated using the proposed cross-attention mechanism.

Given the input graph sequence $\mathcal{G}$, the objective is to learn a function.

$$f(\mathcal{G}; \Theta) \rightarrow \hat{Y}, \quad (2)$$

where Θ denotes the learnable parameters and $\hat{Y}$ represents the predicted disinformation class. For binary classification, the prediction can be expressed as

$$\hat{y}_i = \arg \max_{c \in \{0,1\}} p(y_i = c \mid \mathcal{G}), \quad (3)$$

where $y_i = 1$ denotes disinformation and $y_i = 0$ denotes non-disinformation, subject to the ground-truth labeling procedure described in Section 2.2.

The formulation explicitly distinguishes the structural and temporal dimensions of the problem. The graph component captures the relational dependencies among users and interactions, whereas the temporal component captures changes in these representations across successive observations. The cross-attention module is then used to learn relevance-aware interactions between the spatial and temporal representations before the final classification stage.

2.2. Dataset Construction and Preprocessing

The dataset used in this study consists of social-media interactions related to the South China Sea issue. The resulting dataset contains 72,184 tweets generated by 18,932 users and 96,517 interaction edges. Users are represented as nodes, while interactions among users are represented as directed edges. The interaction structure includes retweets, replies, and mentions, allowing the dataset to represent the relational characteristics of information dissemination. Prior to model construction, the collected data were organized according to their temporal information to preserve the chronological characteristics of information propagation. Tweet records were associated with their corresponding users and interaction relationships, after which the interaction data were transformed into temporal graph representations. Each node was associated with the available feature representation, while the edges encoded the interaction relationships among users. The preprocessing procedure consisted of data consistency checking, removal or handling of invalid records, construction of user–interaction relationships, temporal ordering, and transformation of the resulting records into graph-based representations. The final graph representation was subsequently used as the input for the spatial and temporal encoders.

2.3. Dynamic Graph Construction

To represent the temporal evolution of social-media interactions, the complete interaction network is divided into a sequence of time-indexed graph snapshots. At each time step t , the graph is represented as:

$$G_t = (V_t, E_t, X_t), \quad (4)$$

where V_t represents the active nodes, E_t represents the interactions observed during the corresponding temporal interval, and X_t represents the node features. The temporal graph sequence is therefore defined as:

$$\mathcal{G} = \{G_1, G_2, \dots, G_T\}. \quad (5)$$

For each graph snapshot, an adjacency matrix A_t is constructed to encode the observed interactions. Self-loops are subsequently added to preserve the representation of each node during graph convolution:

$$\tilde{A}_t = A_t + I. \quad (6)$$

The corresponding degree matrix is defined as:

$$\tilde{D}_{t,ii} = \sum_j \tilde{A}_{t,ij}. \quad (7)$$

The normalized adjacency matrix is then obtained as:

$$\hat{A}_t = \tilde{D}_t^{-1/2} \tilde{A}_t \tilde{D}_t^{-1/2}. \quad (8)$$

This formulation enables the spatial encoder to process the network structure at each temporal observation while preserving the chronological order required by the subsequent temporal encoder.

2.4. Overall Architecture

The proposed architecture consists of three principal components: (1) a spatial encoder based on GCN, (2) a temporal encoder based on LSTM, and (3) a cross-attention fusion module. For each temporal snapshot G_t , the GCN transforms the graph structure and node features into a spatial representation H_t^S . The sequence of spatial representations is then processed by the LSTM to obtain temporal

representations H_t^T . The proposed cross-attention mechanism subsequently establishes an adaptive interaction between the spatial and temporal representations. The resulting fused representation is passed to the prediction layer to generate the final disinformation classification. The overall computation can be summarized as:

$$\begin{aligned} H_t^S &= \text{GCN}(X_t, A_t), \\ H_{1:T}^T &= \text{LSTM}(H_1^S, H_2^S, \dots, H_T^S), \\ H^F &= \text{CrossAttention}(H^S, H^T), \end{aligned} \quad (9)$$

2.5. Spatial Representation Using GCN

The spatial encoder employs a Graph Convolutional Network to capture structural dependencies among interconnected users. For each temporal graph snapshot G_t , the normalized adjacency matrix $\hat{A}_t$ is used to propagate information across neighboring nodes. The graph convolution at layer l is formulated as:

$$H_t^{(l+1)} = \sigma(\hat{A}_t H_t^{(l)} W^{(l)}), \quad (10)$$

where $H_t^{(l)}$ is the node representation at layer l , $W^{(l)}$ is a trainable weight matrix, and $\sigma(\cdot)$ denotes the nonlinear activation function. The initial representation is defined as:

$$H_t^{(0)} = X_t. \quad (11)$$

After L graph-convolution layers, the spatial representation is obtained as:

$$H_t^S = H_t^{(L)}. \quad (12)$$

Thus, H_t^S contains information derived from both the individual node features and the structural dependencies encoded by the interaction network at time t .

2.6. Temporal Representation Using LSTM

While the GCN captures structural dependencies at each temporal observation, the temporal evolution of these representations is modeled using an LSTM. Given the sequence of spatial representations:

$$\{H_1^S, H_2^S, \dots, H_T^S\}, \quad (13)$$

the LSTM updates its hidden and memory states at each time step. The standard LSTM formulation is expressed as:

$$\begin{aligned} f_t &= \sigma(W_f H_t^S + U_f h_{t-1} + b_f), \\ i_t &= \sigma(W_i H_t^S + U_i h_{t-1} + b_i), \\ \tilde{C}_t &= \tanh(W_c H_t^S + U_c h_{t-1} + b_c), \\ C_t &= f_t \odot C_{t-1} + i_t \odot \tilde{C}_t, \\ o_t &= \sigma(W_o H_t^S + U_o h_{t-1} + b_o), \\ h_t &= o_t \odot \tanh(C_t), \end{aligned} \quad (14)$$

where f_t , i_t , and o_t denote the forget, input, and output gates, respectively; C_t represents the memory state; h_t represents the hidden state; and $\odot$ denotes element-wise multiplication. The resulting hidden representations form the temporal representation:

2.7. Cross-Attention Fusion Mechanism

The central component of the proposed framework is the cross-attention fusion mechanism, which is designed to establish an adaptive interaction between spatial and temporal representations. Unlike direct concatenation, which combines representations without explicitly modeling their relevance, cross-attention enables one representation to selectively attend to informative components of another representation. Let H^S denote the spatial representation generated by the GCN and H^T denote the temporal representation generated by the LSTM. The query, key, and value representations are obtained through learnable linear projections:

$$\begin{aligned} Q &= H^T W_Q, \\ K &= H^S W_K, \\ V &= H^S W_V, \end{aligned} \quad (15)$$

where W_Q , W_K , and W_V are trainable projection matrices.

The attention weights are calculated as:

$$A = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right), \quad (16)$$

where d_k represents the dimensionality of the key representation. The cross-attended representation is then obtained by:

$$H^{CA} = AV. \quad (17)$$

In this formulation, the temporal representation acts as the query while the spatial representation provides the keys and values. Therefore, the model learns which structural information is relevant to the temporal state being analyzed. This mechanism provides an adaptive interaction between temporal evolution and graph structure rather than treating the two representations as independent features.

2.8. Residual Connection and Layer Normalization

To improve representation stability during feature fusion, the cross-attended representation is combined with the original temporal representation through a residual connection:

$$H^R = H^T + H^{CA}. \quad (18)$$

The resulting representation is subsequently normalized using layer normalization:

$$H^F = \text{LayerNorm}(H^R). \quad (19)$$

The residual connection preserves information from the original temporal representation while allowing the cross-attention output to provide additional spatial information. Layer normalization is applied to stabilize the resulting feature distribution and facilitate optimization.

2.9. Prediction Layer

The fused representation H^F is passed to a fully connected classification layer to generate the final class probabilities. For binary disinformation classification, the output probability is expressed as:

$$\hat{y} = \text{softmax}(W_o H^F + b_o), \quad (20)$$

where W_o and b_o denote the trainable output parameters. The predicted class corresponds to the class with the highest probability.

2.10. Training and Optimization

The proposed model is trained in a supervised learning setting using the labeled samples in the training set. The training process minimizes the classification loss between the predicted and ground-truth labels. For binary classification, the cross-entropy loss is defined as:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)], \quad (21)$$

where N denotes the number of training instances, y_i is the ground-truth label, and $\hat{y}_i$ is the predicted probability.

2.11 Experimental Protocol and Baselines

The dataset is divided into training, validation, and testing subsets using the predefined experimental protocol. The same data partition and evaluation procedure are applied to all baseline models to ensure a fair comparison. The baseline models include CNN, LSTM, GNN, BERT, and GAT-BERT, while the proposed model combines GNN, LSTM, and cross-attention. Each baseline is trained using its corresponding input representation while maintaining comparable training and evaluation conditions. Model-specific hyperparameters are selected using the training and validation sets and are not adjusted using the test set.

2.12 Evaluation Metrics

The performance of each model is evaluated using accuracy, precision, recall, F1-score, macro-F1, weighted-F1, balanced accuracy, Matthews correlation coefficient (MCC), and area under the receiver operating characteristic curve (AUC). These metrics are selected to provide a comprehensive assessment of classification performance, particularly in the presence of potential class imbalance.

$$\begin{aligned}
Accuracy &= \frac{TP + TN}{TP + TN + FP + FN} \\
Precision &= \frac{TP}{TP + FP} \\
Recall &= \frac{TP}{TP + FN} \\
F1 &= 2 \frac{Precision \times Recall}{Precision + Recall} \\
MCC &= \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}
\end{aligned} \tag{22}$$

3. RESULTS AND DISCUSSIONS

3.1 Experimental Results

The experimental evaluation was conducted to assess the effectiveness of the proposed GNN-LSTM framework with cross-attention for disinformation detection in social-media interactions associated with the South China Sea. The evaluation focuses on four primary classification metrics, namely accuracy, precision, recall, and F1-score. The proposed model is compared with CNN, LSTM, GNN, BERT, and GAT-BERT to evaluate its performance relative to conventional deep-learning, graph-based, and transformer-based approaches. The dataset used in the experiment consists of 72,184 tweets generated by 18,932 users and connected through 96,517 interaction edges. The interaction network represents retweets, replies, and mentions, allowing the experimental setting to incorporate both relational and temporal characteristics of information propagation. The comparative results demonstrate that the proposed architecture achieves the highest performance among the evaluated models. The proposed GNN-LSTM with cross-attention obtains an accuracy of 95.84%, precision of 95.30%, recall of 95.60%, and F1-score of 95.45%. The detailed comparison is presented in Table 1.

Table 1. Performance Comparison of Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
CNN	82.45	81.90	82.10	81.95
LSTM	84.12	83.75	84.00	83.87
GNN	86.78	86.20	86.50	86.35
BERT	89.65	89.10	89.40	89.25
GAT-BERT	93.21	92.80	93.00	92.90
GNN-LSTM + Cross-Attention	95.84	95.30	95.60	95.45

Table 1 shows a progressive improvement in model performance, from CNN (82.45%) and LSTM (84.12%) to GNN (86.78%), BERT (89.65%), and GAT-BERT (93.21%). The proposed GNN-LSTM with cross-attention achieves the highest accuracy of 95.84%, representing gains of 2.63 percentage points in accuracy, 2.50 in precision, 2.60 in recall, and 2.55 in F1-score over GAT-BERT. These results indicate that integrating graph-based structural dependencies, temporal dynamics, and cross-attention enables more effective spatio-temporal representation than relying on textual, structural, or sequential information alone.

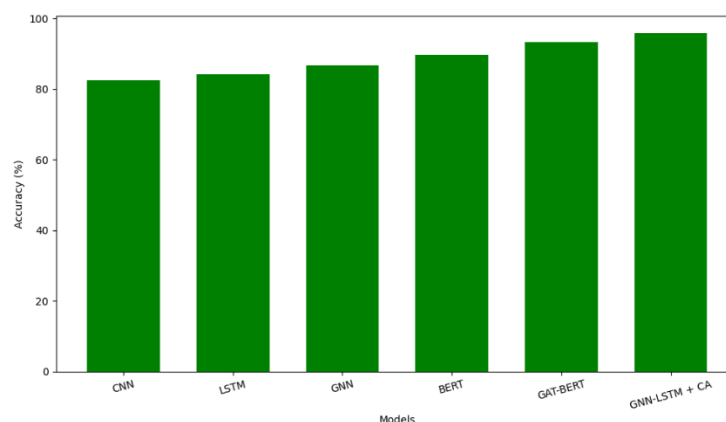


Figure 1. Accuracy Comparison of Models

3.2 Performance Improvement over the Strongest Baseline

To provide a more detailed assessment of the performance advantage, the proposed model is compared with GAT-BERT, which achieves the highest performance among the baseline approaches. Rather than reporting only the absolute performance values, the analysis calculates both absolute improvement in percentage points and relative improvement.

Table 2. Performance Improvement of the Proposed Model over GAT-BERT

Metric	GAT-BERT (%)	Proposed (%)	Absolute Improvement (pp)	Relative Improvement (%)
Accuracy	93.21	95.84	2.63	2.82
Precision	92.80	95.30	2.50	2.69
Recall	93.00	95.60	2.60	2.80
F1-score	92.90	95.45	2.55	2.74

GAT-BERT achieves an accuracy of 93.21% and an F1-score of 92.90%, while the proposed model improves accuracy by 2.63 percentage points and F1-score by 2.55 percentage points, with additional gains of 2.50 and 2.60 percentage points in precision and recall, respectively. The consistent improvement across all four metrics indicates that the proposed model provides more balanced classification performance rather than improving accuracy alone.

3.3 Ablation Study

An ablation study was conducted to determine whether the performance improvement of the proposed architecture is associated with the individual spatial and temporal components and, more importantly, with the cross-attention fusion mechanism. The analysis compares the standalone LSTM and GNN configurations with the combined GNN-LSTM architecture and the final GNN-LSTM architecture incorporating cross-attention.

Table 3. Ablation Study of the Proposed Architecture

Model Configuration	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
LSTM	84.12	83.75	84.00	83.87
GNN	86.78	86.20	86.50	86.35
GNN + LSTM	92.10	91.80	92.00	91.85
GNN-LSTM + Cross-Attention	95.84	95.30	95.60	95.45

The ablation results show progressive performance improvements as spatial, temporal, and adaptive fusion components are added. The GNN-only model achieves 86.78% accuracy and 86.35% F1-score, while incorporating LSTM increases these values to 92.10% and 91.85%, representing gains of 5.32 and

5.50 percentage points, respectively. Adding cross-attention further improves accuracy to 95.84% and F1-score to 95.45%, with additional gains of 3.74 and 3.60 percentage points. These results indicate that cross-attention provides complementary benefits by adaptively integrating spatial and temporal representations beyond the independent GNN and LSTM components.

3.4 Incremental Contribution of Model Components

To further clarify the contribution of each component, the incremental performance gains are calculated relative to the immediately preceding architecture. This analysis is particularly important because the principal methodological contribution of the study is the cross-attention mechanism rather than the simple combination of GNN and LSTM.

Table 4. Incremental Contribution of Model Components

Model Configuration	Accuracy (%)	Accuracy Gain (pp)	F1-score (%)	F1 Gain (pp)
GNN	86.78	–	86.35	–
GNN + LSTM	92.10	+5.32	91.85	+5.50
GNN + LSTM + Cross-Attention	95.84	+3.74	95.45	+3.60

Table 4 shows that integrating LSTM into GNN improves accuracy and F1-score by 5.32 and 5.50 percentage points, respectively, indicating the complementary contribution of temporal information beyond spatial graph representations. The subsequent addition of cross-attention provides further gains of 3.74 percentage points in accuracy and 3.60 percentage points in F1-score, demonstrating that adaptive spatial-temporal fusion contributes additional performance beyond simply combining GNN and LSTM.

3.5 ROC and AUC Analysis

The discriminative performance of the proposed GNN–LSTM model with cross-attention was further examined using Receiver Operating Characteristic (ROC) analysis. This evaluation assesses the model's ability to distinguish between disinformation and non-disinformation instances across different classification thresholds by analyzing the relationship between the True Positive Rate (TPR) and False Positive Rate (FPR). The resulting ROC curve is presented in Figure 3.

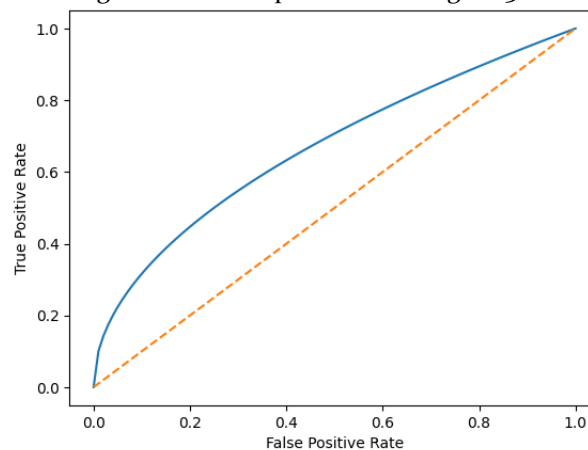


Figure 3. ROC Curve

Figure 3 presents the ROC curve of the proposed GNN–LSTM model with cross-attention, illustrating the trade-off between the True Positive Rate (TPR) and False Positive Rate (FPR) across different classification thresholds. The ROC curve remains consistently above the diagonal reference line, indicating that the proposed model provides discriminative capability beyond random classification. The curve shows a relatively rapid increase in TPR at lower FPR values, demonstrating the model's ability to identify a considerable proportion of disinformation instances while limiting false-positive predictions. Overall, the ROC analysis supports the effectiveness of the proposed model in

distinguishing between disinformation and non-disinformation instances across varying decision thresholds.

3.6 Discussion

The combined findings demonstrate that disinformation propagation in social-media environments can be more effectively characterized by jointly modeling relational structure and temporal evolution. The progressive improvement from CNN and LSTM to GNN and GAT-BERT indicates that textual, sequential, and relational representations capture complementary characteristics of the information environment. The proposed GNN-LSTM architecture further integrates these dimensions, where GNN captures dependencies among interacting entities and LSTM models their temporal evolution. The ablation results provide clear empirical support for this integration: incorporating LSTM into GNN improves accuracy by 5.32 percentage points and F1-score by 5.50 percentage points, while adding cross-attention produces further gains of 3.74 and 3.60 percentage points, respectively. These results indicate that the contribution of the proposed method is not merely the combination of GNN and LSTM, but the adaptive fusion of spatial and temporal representations. The cross-attention mechanism enables the model to dynamically identify relevant structural information according to the current temporal state. This capability is particularly relevant in dynamic social-media environments, where interaction patterns and the relevance of particular entities may change as an information narrative develops. Unlike simple feature concatenation, cross-attention explicitly models the interaction between temporal and spatial representations, producing a context-dependent fused representation. The improvement from 92.10% to 95.84% accuracy and from 91.85% to 95.45% F1-score after introducing cross-attention provides quantitative evidence for this contribution. From a defense-oriented perspective, the proposed framework offers a potential decision-support approach for analyzing disinformation propagation in geopolitical information environments, including issues related to the South China Sea. By combining interaction structures with temporal propagation patterns, the model may assist analysts in identifying emerging patterns that are difficult to capture from textual information alone. However, the framework should be regarded as an analytical support tool rather than an autonomous mechanism for determining user intent, responsibility, or identity, since its predictions represent statistical patterns learned from the available data and require contextual and domain-specific interpretation. Several limitations should also be considered. The evaluation is based on a South China Sea-related dataset, so generalization to other geopolitical contexts, platforms, and information environments remains to be established. Model performance may also be influenced by the temporal resolution, interaction characteristics, and changes in user behavior or platform dynamics. In addition, cross-attention weights should not be interpreted as causal explanations because learned attention indicates model relevance rather than causal relationships. Future research should therefore validate the framework using independent and cross-domain datasets, repeated experimental runs, confidence intervals, and statistical significance testing to strengthen evidence of robustness and generalizability.

4. CONCLUSION

This study proposed a hybrid deep learning framework that integrates Graph Neural Networks (GNN) and Long Short-Term Memory (LSTM) with a cross-attention fusion mechanism for modeling and predicting disinformation propagation in the context of information warfare in the Laut Cina Selatan. The proposed approach was designed to address the limitations of conventional methods by simultaneously capturing structural relationships and temporal evolution within complex information diffusion networks. Experimental results demonstrated that the proposed GNN-LSTM with cross-attention model consistently outperformed all baseline approaches, including CNN, LSTM, GNN, BERT, and GAT-BERT. The model achieved the highest performance with an accuracy of 95.84%, precision of 95.30%, recall of 95.60%, and F1-score of 95.45%. These findings confirm that disinformation propagation is inherently a spatio-temporal phenomenon that requires integrated modeling of network topology and temporal dynamics. Furthermore, the ablation study verified that each architectural component contributed significantly to performance improvement, with the cross-

attention mechanism emerging as the key factor enabling effective interaction between spatial and temporal representations. The ROC analysis further demonstrated the robustness and discriminative capability of the proposed framework, indicating its effectiveness in distinguishing disinformation from non-disinformation instances while maintaining low false positive rates. This capability is particularly important in geopolitically sensitive regions where information manipulation can influence public perception, strategic stability, and national security. From a theoretical perspective, this research contributes to the advancement of intelligent information warfare analytics by introducing a unified spatio-temporal fusion framework based on attention mechanisms. From a practical perspective, the proposed model provides a scalable and robust solution for real-time monitoring and mitigation of disinformation campaigns in maritime conflict environments. Nevertheless, this study has several limitations. The experiments relied primarily on Twitter-based datasets, which may not fully represent cross-platform disinformation ecosystems. In addition, the current framework focuses mainly on textual propagation patterns without incorporating multimodal information such as images, videos, or audio content. Future research may extend this work by integrating multimodal deep learning, transformer-based graph architectures, and real-time streaming analytics to further improve the adaptability and generalization of disinformation detection systems in dynamic cyber-defense environments.

DECLARATIONS

AI USAGE STATEMENT

The authors declare that no artificial intelligence (AI) tools or AI-assisted applications were used in the research, data analysis, manuscript preparation, writing, or editing of this article. All research activities and the preparation of the manuscript were carried out by the authors, who take full responsibility for the originality, accuracy, and integrity of the work.

AUTHOR CONTRIBUTION

Adam Mardamsyah contributed to conceptualization, methodology development, investigation, data curation, formal analysis, software development, validation, and writing the original draft. **Marsetio** contributed to conceptualization, methodology refinement, validation, supervision, and critical revision of the manuscript. **Bambang Irwanto** contributed to data analysis, interpretation of results, validation, visualization, and manuscript review and editing. **Robby Mochammad Taufik** contributed to software implementation, experimental evaluation, data processing, interpretation of findings, and manuscript review and editing. All authors reviewed and approved the final version of the manuscript.

FUNDING STATEMENT

This research was entirely self-funded by the authors. No external funding was received from any government agency, research institution, commercial organization, or other funding body.

CONFLICTING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

- [1] A. Bakirov and I. Suleimenov, "Theoretical Bases of Methods of Counteraction to Modern Forms of Information Warfare," *Computers*, vol. 14, no. 10, p. 410, 2025, doi: <https://doi.org/10.3390/computers14100410>.
- [2] C. Ivan, I. Chiru, and R. Arcos, "A whole of society intelligence approach: critical reassessment of the tools and means used to counter information warfare in the digital age," *Intell. Natl. Secur.*, vol. 36, no. 4, pp. 495–511, 2021, doi: <https://doi.org/10.1080/02684527.2021.1893072>.
- [3] B. Huang, "Navigating digital divide: exploring the influence of ideological and political education on cyber security and digital literacy amid information warfare," *Curr. Psychol.*, vol. 43, no. 28, pp. 23815–23836, 2024, doi: <https://doi.org/10.1007/s12144-024-06106-1>.

- [4] R. Arcos and H. Smith, "Digital communication and hybrid threats," *Rev. ICONO 14. Rev. científica Comun. y Tecnol. emergentes*, vol. 19, no. 1, pp. 1–14, 2021, doi: <https://doi.org/10.7195/ri14.v19i1.1662>.
- [5] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science (80-.)*, vol. 359, no. 6380, pp. 1146–1151, 2018, doi: [10.1126/science.aap9559](https://doi.org/10.1126/science.aap9559).
- [6] Z. Jin, J. Cao, Y. Zhang, and J. Luo, "News verification by exploiting conflicting social viewpoints in microblogs," in *Proceedings of the AAAI conference on artificial intelligence*, 2016, vol. 30, no. 1, doi: <https://doi.org/10.1609/aaai.v30i1.10382>.
- [7] D. I. Sukmawan and R. Pedrason, "Assertively Hedging? Indonesia's Motives and Practices in the South China Sea," *Intermestic J. Int. Stud.*, vol. 7, no. 2, pp. 701–720, 2023.
- [8] C. Paul and M. Matthews, "The Russian 'firehose of falsehood' propaganda model," *Rand Corp.*, vol. 2, no. 7, pp. 1–10, 2016.
- [9] C. Paul and M. Matthews, "Defending against Russian propaganda," *SAGE Handb. Propag.*, pp. 286–302, 2020.
- [10] H. R. Saeidnia, E. Hosseini, B. Lund, M. A. Tehrani, S. Zaker, and S. Molaei, "Artificial intelligence in the battle against disinformation and misinformation: a systematic review of challenges and approaches," *Knowl. Inf. Syst.*, vol. 67, no. 4, pp. 3139–3158, 2025, doi: <https://doi.org/10.1007/s10115-024-02337-7>.
- [11] M. R. Islam, S. Liu, X. Wang, and G. Xu, "Deep learning for misinformation detection on online social networks: a survey and new perspectives," *Soc. Netw. Anal. Min.*, vol. 10, no. 1, p. 82, 2020, doi: [10.1007/s13278-020-00696-x](https://doi.org/10.1007/s13278-020-00696-x).
- [12] K. Alahmadi, S. Alharbi, J. Chen, and X. Wang, "Generalizing sentiment analysis: a review of progress, challenges, and emerging directions," *Soc. Netw. Anal. Min.*, vol. 15, no. 1, p. 45, 2025, doi: <https://doi.org/10.1007/s13278-025-01461-8>.
- [13] P. Lops, D. Jannach, C. Musto, T. Bogers, and M. Koolen, "Trends in content-based recommendation: Preface to the special issue on Recommender systems based on rich item descriptions," *User Model. User-adapt. Interact.*, vol. 29, no. 2, pp. 239–249, 2019, doi: <https://doi.org/10.1007/s11257-019-09231-w>.
- [14] D. Nath and S. K. Dwivedi, "Aspect-based sentiment analysis: approaches, applications, challenges and trends," *Knowl. Inf. Syst.*, vol. 66, no. 12, pp. 7261–7303, 2024, doi: <https://doi.org/10.1007/s41060-017-0082-x>.
- [15] B. Khemani, S. Patil, K. Kotecha, and S. Tanwar, "A review of graph neural networks: concepts, architectures, techniques, challenges, datasets, applications, and future directions," *J. Big Data*, vol. 11, no. 1, p. 18, 2024, doi: [10.1186/s40537-023-00876-4](https://doi.org/10.1186/s40537-023-00876-4).
- [16] V. Ponzi and C. Napoli, "Graph Neural Networks: Architectures, Applications, and Future Directions," *IEEE Access*, vol. 13, pp. 62870–62891, 2025, doi: [10.1109/ACCESS.2025.3558752](https://doi.org/10.1109/ACCESS.2025.3558752).
- [17] H. Rama Moorthy *et al.*, "Dual stream graph augmented transformer model integrating BERT and GNNs for context aware fake news detection," *Sci. Rep.*, vol. 15, no. 1, p. 25436, 2025.
- [18] L. Selvam, E. S. Vinothkumar, R. S. Krishnan, G. V. Rajkumar, J. R. F. Raj, and P. S. R. Malar, "A Unified Deep Learning Model for Fake Account Identification Using Transformer-Based NLP and Graph Neural Networks," in *2025 International Conference on Inventive Computation Technologies (ICICT)*, 2025, pp. 1033–1040.
- [19] B. Lakzaei, M. Haghiri Chehreghani, and A. Bagheri, "Disinformation detection using graph neural networks: a survey," *Artif. Intell. Rev.*, vol. 57, no. 3, p. 52, 2024.
- [20] S. Kuntur *et al.*, "Graph Neural Networks for Misinformation Detection: Performance-Efficiency Trade-offs," *arXiv Prepr. arXiv2604.08131*, 2026.
- [21] P. Gupta and P. K. Gupta, "Performance analysis of gcnn, gnn, and gat models with differentiable pooling for detection of fake news," in *2024 3rd Edition of IEEE Delhi Section Flagship Conference (DELCON)*, 2024, pp. 1–6.
- [22] A. Sankar, Y. Wu, L. Gou, W. Zhang, and H. Yang, "DySAT: Deep Neural Representation Learning on Dynamic Graphs via Self-Attention Networks," pp. 519–527, 2020, doi: <https://doi.org/10.1145/3336191.3371845>.
- [23] A. Pareja *et al.*, "EvolveGCN: Evolving Graph Convolutional Networks for Dynamic Graphs," 2018, doi: <https://doi.org/10.1609/aaai.v34i04.5984>.
- [24] Y. Zhong and C. Huang, "A dynamic graph representation learning based on temporal graph transformer," *Alexandria Eng. J.*, vol. 63, pp. 359–369, 2023, doi: <https://doi.org/10.1016/j.aej.2022.08.010>.
- [25] M. Krichen and A. Mihoub, "Long short-term memory networks: A comprehensive survey," *AI*, vol. 6, no. 9, p. 215, 2025, doi: <https://doi.org/10.3390/ai6090215>.
- [26] M. Rithani, R. P. Kumar, and S. Doss, "A review on big data based on deep neural network approaches," *Artif. Intell. Rev.*, vol. 56, no. 12, pp. 14765–14801, 2023, doi: <https://doi.org/10.1007/s10462-023-10512-5>.

- [27] J. Manurung, P. Sihombing, M. A. Budiman, and Sawaluddin, "Dynamic Rumor Control in Social Networks Using Temporal Graph Neural Networks," in *2023 IEEE International Conference of Computer Science and Information Technology: The Role of Artificial Intelligence Technology in Human and Computer Interactions in the Industrial Era 5.0, ICOSNIKOM 2023*, 2023, pp. 1–5. doi: 10.1109/ICoSNIKOM60230.2023.10364382.
- [28] J. Wang, X. Wang, and A. Yu, "Tackling misinformation in mobile social networks a BERT-LSTM approach for enhancing digital literacy," *Sci. Rep.*, vol. 15, no. 1, p. 1118, 2025, doi: <https://doi.org/10.1038/s41598-025-85308-4>.
- [29] L. Abualigah, Y. Y. Al-Ajlouni, M. S. Daoud, M. Altalhi, and H. Migdady, "Fake news detection using recurrent neural network based on bidirectional LSTM and GloVe," *Soc. Netw. Anal. Min.*, vol. 14, no. 1, p. 40, 2024, doi: <https://doi.org/10.1007/s13278-024-01198-w>.
- [30] G. Loukides and R. Gwadera, "Preventing the diffusion of information to vulnerable users while preserving pagerank," *Int. J. Data Sci. Anal.*, vol. 5, no. 1, pp. 19–39, 2018, doi: <https://doi.org/10.1007/s41060-017-0082-x>.
- [31] J.-X. Xia, W.-W. Jia, P. Li, and X.-K. Xu, "MSTD: A Joint Structural–Temporal Framework for Fake News Detection," *IEEE Trans. Comput. Soc. Syst.*, 2026, doi: 0.1109/TCSS.2026.3674080.