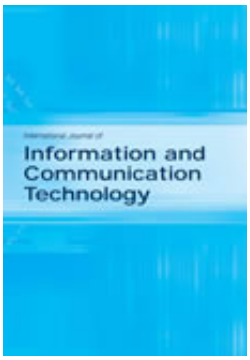
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Information and Communication Technology

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

Social network news dissemination and detection model based on multi-scale information

Yanli Zhang

DOI: 10.1504/IJICT.2026.10075952

Article History:

Received:	10 October 2025
Last revised:	25 November 2025
Accepted:	01 December 2025
Published online:	06 February 2026

Social network news dissemination and detection model based on multi-scale information

Yanli Zhang

Publicity Department,
Henan Open University,
Zhengzhou, 450046, China
Email: zylhenan@hotmail.com

Abstract: At a critical stage of social network information evolution, the rapid spread of false news, emotional content, and rumours poses significant challenges to information governance. To address this issue, a multi-scale social network news dissemination and detection model is proposed. The model integrates multi-scale feature extraction, cascaded convolutional networks, and cross-modal information modelling to enhance feature representation and propagation pattern capture. Experimental results show that introducing macro-micro dual-scale modelling and gated fusion improves the F1 score to 0.895 and reduces the mean absolute percentage error to 8.9%, representing gains of 7.8 and 3.4 percentage points over single-scale baselines ($p < 0.01$). Across diverse communication scenarios, the model consistently outperforms comparison methods, achieving macro-F1 scores of 86%-91% and micro-F1 scores of 88%–92%. With an average detection delay of approximately 12 ms, the model balances real-time performance and robustness, demonstrating effectiveness and stability for multi-scenario news detection.

Keywords: news detection; multi-scale feature extraction; cascaded convolutional network; cross-modal information; gating mechanism.

Reference to this paper should be made as follows: Zhang, Y. (2026) 'Social network news dissemination and detection model based on multi-scale information', *Int. J. Information and Communication Technology*, Vol. 27, No. 6, pp.1–26.

Biographical notes: Yanli Zhang graduated from Zhengzhou University with a Master's in Journalism and Communication. Currently, she teaches at Henan Open University as a lecturer and holds the title of 'Dual-qualified' instructor in Intermediate Network Journalism and Communication. Her research focuses on journalism and communication studies. She has led and completed five research projects funded by the Henan Social Sciences Association and the Municipal Social Sciences Association, and has published over ten academic papers.

1 Overview

In recent years, social networks have become an important platform for information dissemination, but their openness and immediacy have also exacerbated the rapid spread of false news, rumours, and other information (Wu et al., 2025). This type of information not only misleads public perception, but may also trigger social panic and loss of control

over public opinion. How to efficiently and accurately identify fake news and suppress its spread has become a critical research topic in the fields of natural language processing and social network analysis (Wang and Zhang, 2024). Existing research has proposed various methods in news detection, including deep convolutional neural networks, graph neural networks, and cross-modal fusion models. In terms of traditional machine learning methods, early research on false news detection mainly relied on text surface features and simple statistical methods (Wang et al., 2024). M A Ilyas et al. used various feature extraction methods, including counting vectorisers, bag of words, word representation global vectors, word-to-vector and word frequency inverse document frequency, as well as feature selection techniques such as information gain, chi square test, principal component analysis, and document frequency. The accuracy and robustness of fake news detection were significantly improved through ensemble learning (Ilyas et al., 2024). Q Liu et al. (2024) proposed a new dual-adversarial learning method to address the issue of bias between news and evidence content. The core innovation of this framework lies in the simultaneous construction of de biased discriminators on both the news and evidence sides, both of which undergo adversarial training targeting true and false news labels. By reverse optimising these two discriminators, this method could effectively eliminate potential biases in news content and evidence materials. Dua et al. (2023) developed the interpretable fake linguistic analysis and semantic heuristics (I-FLASH) model. This model has dual capabilities of detection and attribution. It can not only distinguish the authenticity of news, but also automatically generate the basis for determining whether it is true or false by analysing features such as content and source, thus providing users with transparent decision explanations.

With the breakthrough of deep learning in natural language processing and multimodal analysis, researchers have begun to explore the use of deeper semantic representations and structural modelling to enhance detection capabilities (Cao et al., 2025). Birunda et al. (2024) proposed a fake news detection method based on honey badger optimisation algorithm and lightweight convolutional random forest. This method achieves feature minimisation while ensuring high accuracy through three steps: data preprocessing, feature selection, and classification. The experimental results show that the model performs well in multiple indicators such as accuracy, verifying its effectiveness and advantages in fake news detection tasks. Zhai et al. (2023) proposed a hybrid model of multi-scale CNN and long short term memory (LSTM) by combining the advantages of CNN and recurrent neural network (RNN). The model first extracted local text features through multi-scale CNN, then captured contextual dependencies using LSTM, and finally fused the generated feature vectors into a softmax classifier. Experiments showed that the hybrid model outperformed traditional CNN, LSTM and other single models in text classification tasks, showing significant advantages. Suryawanshi et al. (2024) proposed an incremental ensemble neural network model with continuous learning capability for fake news detection. This model optimised classifier combinations through a performance-based dynamic pruning mechanism and monitored concept drift in real-time to automatically adjust detection strategies, thus adapting to the dynamic changes in news patterns. The experimental results showed that in a static testing environment, the performance of this model was superior to traditional machine learning methods.

To sum up, although the existing methods have realised good results in a single task, there are still shortcomings in multi scene migration, anti-jamming capability and capture accuracy for different propagation modes. On the one hand, traditional models often

focus on single-scale features and overlook the improvement of detection performance by cross scale information. On the other hand, existing methods exhibit significant performance degradation in high noise and cross-domain environments, limiting their application in real social networks. To address the limitations of the above methods in cross scene migration, diffusion level recognition, and anti-interference ability, further research has been conducted to construct a technology centred on 'multi-scale feature extraction cross level information interaction cascaded convolution modelling'. This technology emphasises starting from the multidimensional characteristics of the propagation chain and establishing a collaborative expression mechanism between different scales of time series and network structure to achieve the unified representation of local semantic patterns and global diffusion laws. Based on this, a multi-scale information driven cascade convolutional networks (CasCN) model for social network news dissemination and detection is proposed, which achieves the collaborative expression of propagation chain structure features and semantic temporal features, aiming to improve the detection accuracy and robustness of the model in multiple scenarios. Cross modal information modelling mainly focuses on the collaborative fusion between text semantic features and propagation structural features, without involving traditional multimodal input forms such as images and audio. The research regards node text content as semantic mode, propagation path and cascade structure as structural mode, and realises cross modal feature interaction and dynamic fusion through multi-scale coding and gating mechanism, so as to improve the adaptability and detection accuracy of the model to different propagation forms. The specific implementation mechanism is as follows.

2 Methodology

2.1 Information cascade model of the improved CasCN based on multi-scale information

In the social media environment, the diffusion process of information often presents complex network cascading characteristics. Different types of news show significant differences in their dissemination process. For example, real news usually relies on a stable social relationship chain to gradually spread, while fake news is more likely to achieve early concentrated outbreaks through a small number of highly active users or opinion leaders, and form widespread dissemination in a short period of time. With the multi-layered interaction and forwarding behaviour of users, these cascading processes not only exhibit rapidly evolving dynamics in the time dimension, but also manifest as diverse and hierarchical propagation paths in the structural dimension. However, traditional detection methods are difficult to effectively capture the dual features of structure + time sequence, so new modelling methods are needed to incorporate social network propagation itself into the detection framework. Therefore, the information cascade model has received widespread attention. This model can abstract the propagation chain into a graph structure and reveal the differences between false information and real information in the propagation trajectory by modelling the relationships and dynamic evolution between different nodes, laying a theoretical and practical foundation for the introduction of subsequent deep learning methods. Among them, CasCNs abstract the information propagation process as an ordered graph structure, and jointly model the structural features and temporal dependencies of propagation nodes

through convolution operations, thereby providing high-quality feature inputs for downstream prediction tasks (Zhu et al., 2022). This mechanism not only captures fine-grained patterns in local neighbourhoods, but also extracts high-order features of the global propagation path through multi-layer convolution accumulation, making the model more capable of pattern recognition when dealing with complex propagation networks. The CasCN structure is shown in Figure 1.

Figure 1 CasCN structure (see online version for colours)

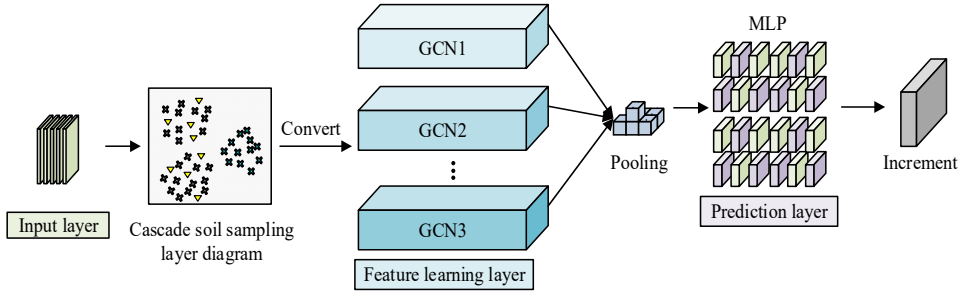
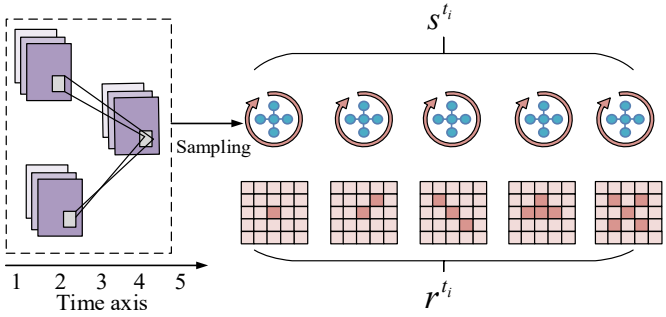


Figure 2 Sampling process of sub cascaded graph (see online version for colours)



In Figure 1, CasCN adopts an end-to-end framework, and the input layer uses an embedding mapping mechanism based on propagation chain sequences to encode text semantic vectors and node interaction features into a 128 dimensional vector; The graph convolutional layer uses a two-layer graph convolutional neural network (GCN) structure to capture first-order and second-order diffusion neighbourhood features, respectively. The activation function uses LeakyReLU and the parameter sharing rate is 0.75 to reduce depth computation redundancy. The feature aggregation stage adopts a weighted sum and residual connection mechanism to suppress gradient decay while ensuring the preservation of diffusion information. The node level output integrates semantic, interaction sequence, and propagation strength features through the fusion layer, and maps them to a 256 dimensional detection vector for subsequent classification prediction. Finally aggregates the features and uses a multi layer perceptron (MLP) for cascading scale prediction (Zhang et al., 2024a). To address the limitations of existing methods, a dynamic sampling method based on timestamps is proposed, which decomposes the original cascaded graph into a sequence of temporal subgraphs S^{t_o} according to node timestamps. The mathematical expression is shown in equation (1) (Tang et al., 2023).

$$S^{t_o} = \left\{ s^{t_i} \middle| t_i \le t_o \right\}, \quad \forall s^{t_i} \in S^{t_o}, t_i \le t_o \tag{1}$$

In equation (1), S^{t_o} represents the sub cascade graph of the deadline t_o, s^{t_i} represents the sub cascade graph generated at time t_i, $t_i \le t_o$ is the constraint condition, indicating that the timestamps of these sub cascade graphs are not later than the deadline t_o. The sampling process is shown in Figure 2 (Huang et al., 2024).

In Figure 2, the cascaded sampling process gradually divides the original propagation chain into multiple sub cascaded graphs according to time constraints. Each sub cascaded graph is represented in the form of an adjacency matrix, where the rows and columns of the matrix correspond to the connection relationships between nodes and edges in the propagation network. During the sampling process, each subgraph starts from the central node and constructs a local propagation structure through neighbourhood aggregation strategy. The adjacency matrix is processed using symmetric normalisation to suppress the bias effect introduced by node degree differences. Multiple cascaded graphs are generated in a time sliding window manner, enabling the model to establish a progressive correlation between micro sequence and macro structure development. From this, the matrix representation of the sub cascade graph sequence can be obtained, as shown in equation (2) (Zeng and Xiang, 2023).

$$R^{t_o} = \left\{ r^{t_1}, ..., r^{t_i} \right\}, t_i \le t_o \tag{2}$$

In equation (2), R^{t_o} represents the adjacency matrix sequence corresponding to the deadline t_o and the sub cascade graph, and r^{t_i} represents the attribute vector associated with s^{t_i}. CasCN utilises graph embedding and a novel cascaded sampling method to effectively integrate spatiotemporal features and achieve good prediction performance. However, its timestamp based sampling strategy has two key drawbacks: firstly, dense timestamps lead to an expansion in the number of subgraphs, significantly increasing the computational burden. The second issue is that the difference between adjacent time point plots is too small, which can easily introduce modelling bias (Preethi and Mamatha, 2023). To this end, a multi-scale graph capsule network (MSGCNet) is proposed to fully capture the propagation characteristics of news across levels and time scales in social networks, and to enhance the robustness and structural awareness of node representations. Compared to traditional graph convolutional networks that rely on fixed order neighbourhood aggregation, MSGCNet constructs a multi-scale structure through multi-stage convolution kernels and dynamic interval sampling in receptive field extension, allowing nodes to simultaneously capture information from different depths and directions, enhancing their perception of long-chain paths. Meanwhile, in terms of parameter sharing mechanism, MSGCNet adopts a cross scale weight collaborative strategy to alleviate the problem of excessive smoothing caused by multi-layer stacking and enhance sensitivity to structural differences. In addition, in terms of feature fusion, MSGCNet introduces position encoding and capsule routing to achieve hierarchical interaction among nodes, categories, and graph levels, effectively modelling the nonlinear dependencies between multi-scale features. MSGCNet has expanded its structural coverage, optimised information transmission efficiency and parameter utilisation, and improved detection accuracy and robustness in complex propagation scenarios. The network consists of two main parts: node embedding layer and multi-level capsule mixing layer. The node embedding layer uses a multiscale graph network (MGN) to learn node representations of sub cascaded graphs through three dimensions: direction,

position, and high-order (Li et al., 2023). MGN reconstructs the convolution kernel as shown in equation (3) (Zhang et al., 2024b).

$$G = s_\alpha \cdot X = \sigma\left[\Big\|_{k \in O}, \Big\|_{\varphi \in \{in,out\}} \left(\hat{R}_\varphi^k X W_\varphi^k \right), PW_p \right] \tag{3}$$

In equation (3), G is the node representation matrix obtained by convolution operation, X is the node feature matrix, and $s_\alpha \cdot X$ represents the convolution operation using the MGN convolution kernel with parameter α applied to the input features. $\|\cdot$ is the feature concatenation operation, $\hat{R}_\varphi^k$ means the normalised k order adjacency matrix, and W_φ^k means the trainable weight matrix in k order and direction φ. P represents the position encoding matrix, and W_p represents the trainable weight matrix corresponding to the position encoding. The asymmetric normalisation form of directional adjacency matrix is shown in equation (4) (Feng et al., 2022).

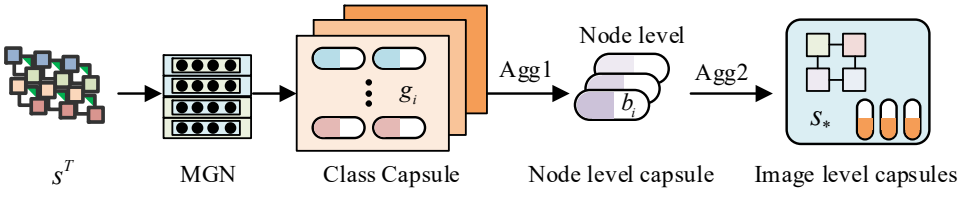
$$\begin{cases} \hat{R}_\varphi = (\overline{A})^{-1} \overline{R}_\varphi \\ \overline{R}_\varphi = R_\varphi + U_N \end{cases} \tag{4}$$

In equation (4), $\overline{A}$ is the degree matrix, $\overline{R}_\varphi$ represents the adjacency matrix after adding self loops in direction φ, U_N represents the identity matrix of $N \times N$, and R_φ is the original adjacency matrix. The position embedding matrix initialised based on position encoding is shown in equation (5).

$$\begin{cases} p_{2d} = \sin\left(\dfrac{u}{10{,}000^{\frac{2d}{d_p}}} \right) \\ p_{2d+1} = \cos\left(\dfrac{u}{10{,}000^{\frac{2d}{d_p}}} \right) \end{cases} \tag{5}$$

In equation (5), p_{2d} and p_{2d+1} are the values of the $2d^{\text{th}}$ and $2d{+}1^{\text{th}}$dimensions of the position vector, respectively. u represents the position information index of the node, d is the index of the embedding dimension, and d_p represents the total dimension size of the position embedding. The node embedding layer encodes multi-scale structures and location information into high-dimensional features, and achieves dynamic routing and multi-level feature fusion through capsule networks to generate robust cascaded graph global representations. This can better capture the diverse path characteristics and semantic differences of social network news in the dissemination process. The multi-level capsule mixing layer includes three types of capsules: class level, node level, and graph level. The structure is shown in Figure 3.

Figure 3 Multi level capsule encoding and aggregation process (see online version for colours)



In Figure 3, firstly, input the cascaded graph s^T into MGN to complete the initial embedding and extraction of node semantics and local structural relationships, with an output dimension of 128. Subsequently, the embedding performs feature pattern classification through a class capsule layer, extracting node features g_i of different orders. Each g_i corresponds to a potential propagation structure type, and vector encoding is used to maintain directional semantics. The first stage feature aggregation (Agg1) utilises a dynamic routing mechanism to calculate node level features b_i, with the goal of identifying key propagation nodes and preserving cross node relationship information. The second stage aggregation (Agg2) maps node level capsules to graph level feature vectors s_*, completing the global integration of multi node features. To capture the dynamic changes of cascading influence, a neural network based on sub cascading graph design is studied, and the attenuation law of influence with interval index is modelled through attention mechanism. The impact on attention is shown in equation (6) (He et al., 2023).

$$a_l = \frac{\exp\left(\left\langle w, \phi \otimes s_l \right\rangle\right)}{\sum_{i=1}^{l} \exp\left(\left\langle w, \phi \otimes s_l \right\rangle\right)} \tag{6}$$

In equation (6), a_l represents the influence attention coefficient of the l^{th} graph level capsule, w means the influence attention weight vector, ϕ is the attenuation factor, and s_l represents the feature representation of the l^{th} graph level capsule. The updated image level capsule is shown in equation (7).

$$s_l' = a_l s_l \tag{7}$$

In equation (7), s_l' represents the updated graph level capsule. Research introduces auxiliary classification tasks in cascade prediction to predict whether the propagation scale exceeds the preset threshold. Class-level capsules are generated through dynamic routing, and the final cascaded representation is obtained by applying a weighted sum operation to the class capsules, as shown in equation (8) (Ren et al., 2023).

$$\begin{cases} w_m = \dfrac{\exp\left(\left\|e_m\right\|\right)}{\sum_{m=1}^{M} \exp\left(\left\|e_m\right\|\right)} \\ E' = \sum_m w_m e_m \end{cases} \tag{8}$$

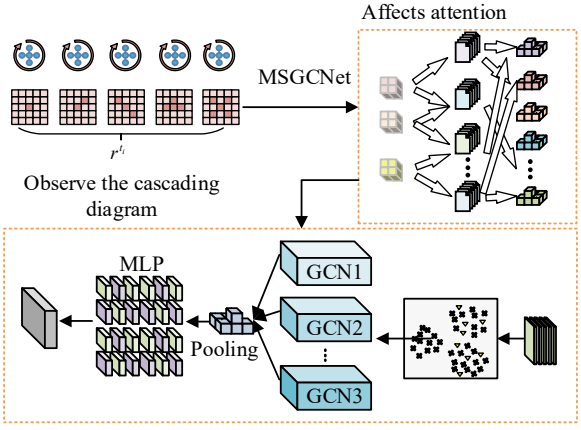
In equation (8), e_m represents the m^{th} capsule, w_m means the weight coefficient of the m^{th} capsule, E' is the weighted global information cascade representation vector, and M is the number of capsules. The definition of the global loss function is denoted in equation (9).

$$L = \frac{1}{b} \sum_{i=1}^{b} \left(\theta \eta_1^i + (1-\theta)\eta_2^i\right) \tag{9}$$

In equation (9), L means the total loss value, b means the batch size, and θ is the weight coefficient used to balance the contributions of η_1^i and η_2^i. η_1^i and η_2^i represent the loss values of the i^{th} sample under the first and second loss terms, respectively. Finally, a multi-scale CasCN (M-CasCN) framework is constructed by combining the sub cascade

graph sampling strategy, MSGCNet, and influence attention mechanism. The influence attention mechanism is used to dynamically adjust the feature contribution of different nodes and their diffusion paths in the propagation cascade modelling process, in order to highlight key propagation nodes with high diffusion influence. The M-CasCN structure is shown in Figure 4.

Figure 4 M-CasCN framework structure (see online version for colours)



In Figure 4, M-CasCN framework hierarchically samples the cascaded graph according to the propagation time period, forming multiple sub cascaded graphs with phased topological structures, and inputs them into MSGCNet for multi-scale graph embedding modelling. Among them, the GCN module adopts a three-layer stacked structure, with convolution kernel orders of k = 1, 2, and 3, used to capture local interaction patterns, intermediate diffusion directions, and high-order semantic propagation dependencies layer by layer. The node embedding dimensions are 128, 192, and 256, respectively. The model introduces an influence attention mechanism to control the significant gain of features in different diffusion stages. The attention weight vector is jointly generated by a trainable parameter matrix and the propagation path influence coefficient, achieving prominent expression of key turning points. Multi scale embedding outputs are pooled and MLP projected to complete detection and prediction, and demonstrate bottom-up evolutionary modelling capabilities at the structural level.

2.2 Construction of cascaded convolutional network based on multi-scale information fusion

The information cascade model based on M-CasCN can effectively capture the local and overall characteristics of the communication path in social network news dissemination scenarios. However, as the propagation chain deepens, single-scale convolution still easily leads to information dilution or noise amplification, especially when dealing with long chains and cross community propagation, the model's discriminative power is still limited. Therefore, to further enhance the model's ability in propagation path representation and multi-level information fusion, a convolutional network based on attention aggregation mechanism is proposed as the propagation encoding component

based on M-CasCN information cascade model, which includes macro and micro levels. The structure is shown in Figure 5 (Guang et al., 2024).

Figure 5 Spread encoding components, (a) macro code (b) micro code (see online version for colours)

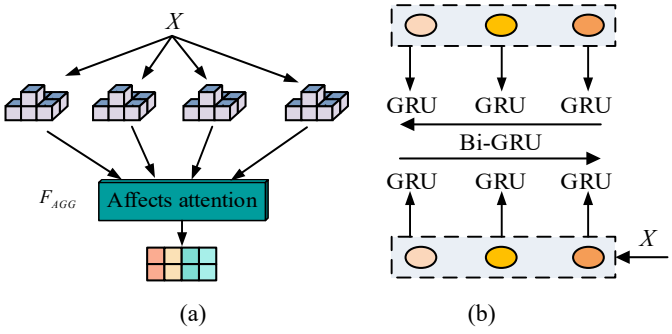


Figure 5(a) shows the macro propagation encoding part. The module takes the cascade graph set X as the input, uses the diffusion depth constraint and time window strategy to generate multiple temporal structure blocks, and selectively enhances different propagation paths based on the influence attention mechanism F_{AGG}. At each level, the blocks first extract diffusion directional features through a structural mapping network, and then combine structural signals such as node forwarding frequency and interaction strength to form multi-scale diffusion embeddings. The attention mechanism dynamically adjusts weights based on the influence of diffusion stages, enabling the model to automatically enhance the expression of key stage features when dealing with explosive or delayed diffusion scenarios. The output encoding matrix contains the structural logic of diffusion patterns in different stages, providing global level guidance information for subsequent convolution and fusion, effectively compensating for the shortcomings of traditional single-scale convolution in characterising the evolution of propagation patterns. Figure 5(b) shows the micro propagation encoding part. The module input is X, and two types of recurrent units, GRU and bidirectional gated recurrent unit (Bi-GRU), are used to capture the forward behaviour development trend and backpropagation dependency relationship, respectively, to solve the problems of information interference and causal misalignment in the sequence. The GRU layer is used to model short-term interaction intensity, such as comment activity or instant forwarding actions; Bi GRU predicts feedback interference and hierarchical delay phenomena in long-chain propagation modes through a bidirectional coupling structure. The model ultimately generates a fusion sequence embedding, which characterises the behaviour transition rules between nodes with time as the main line, and has sensitive responsiveness during the attenuation or re diffusion stage of propagation rhythm. Macro propagation encoding adopts a directed multi-hop graph convolutional network with attention mechanism, and its convolution kernel expression is shown in equation (10) (Zhang et al., 2025).

$$G_{macro} = F_{AGG}\left[\sigma\left(\gamma^{k} X W_{k}\right)_{k\in K}\right] \tag{10}$$

In equation (10), G_{macro} represents the node representation matrix after macro propagation encoding convolution, γ represents the normalised Laplacian operator, and σ is the nonlinear activation function. The study adopts an adaptive order attention mechanism to

dynamically adjust the message transmission distance based on node characteristics, and designs order attention weights for each node in the propagation graph. The calculation is shown in equation (11) (Narayanan et al., 2023).

$$a_u^k = \frac{\exp\left(\left\langle w_u, \tanh\left(W_u g_u^k + b_u\right)\right\rangle\right)}{\sum_{k=1}^{K} \exp\left(\left\langle w_u, \tanh\left(W_u g_u^k + b_u\right)\right\rangle\right)} \tag{11}$$

In equation (11), a_u^k represents the importance weight of node u on its k^{th} hop neighbour features, w_u represents the attention weight vector of node u, g_u^k represents the feature representation vector of node u after aggregation of neighbours in the k^{th} hop, W_u is the weight matrix used for linear transformation of g_u^k, b_u is the trainable bias vector of node u, and tanh(·)is the hyperbolic tangent activation function. The micro propagation encoding adopts the Bi-GRU architecture and memorises the propagation temporal characteristics through a hidden state update mechanism, as shown in equation (12) (Abhilash et al., 2024).

$$G_{micro} = \text{Bi-GRU}\left(x_j, G_{micro}^{j-1}\right), \leftrightarrow G_{micro}^{j} \in R^{F^{micro}} \tag{12}$$

In equation (12), G_{micro} represents the bidirectional GRU hidden state vector, x_j means the input feature vector of the j^{th} time step, and F^{micro} represents the dimension of the hidden state vector in the micro propagation encoding module. The study designed a gating fusion mechanism to dynamically balance the importance of macro and micro encoding features, to achieve adaptive feature fusion, as shown in equation (13) (Yin et al., 2024).

$$\begin{cases} GRU = \sigma\left(W_{gate}^1 G_{Macro} + W_{gate}^2 G_{Micro} + b_{gate}\right) \\ G_{agg} = GRU \otimes G_{Macro} + (1 - GRU) \otimes G_{Micro} \end{cases} \tag{13}$$

In equation (13), GRU is the gating matrix, W_{gate}^1 and W_{gate}^2 correspond to trainable weight matrices for macro and micro features, b_{gate} represents the gating bias vector, and G_{agg} is the final fused feature matrix. Using attention mechanism to weight and aggregate the row features of G_{agg}, the final news representation is constructed as shown in equation (14).

$$G_{news} = \sum_{j=1}^{N} a_i G_{agg} \tag{14}$$

In equation (14), G_{news} represents news representation. After obtaining G_{news}, it is input into an MLP layer with softmax, and the predicted result is shown in equation (15).

$$\hat{y} = \text{softmax}\left(\text{MLP} \cdot G_{news}\right) \tag{15}$$

In equation (15), $\hat{y}$ represents the predicted result. To optimise the performance of news detection, the study introduces the Knowledge Distillation (KD) mechanism, which guides student model training by transferring implicit knowledge from the teacher model (Gou et al., 2022). The definition of the softmax function for introducing temperature regulation is shown in equation (16) (Ma et al., 2024).

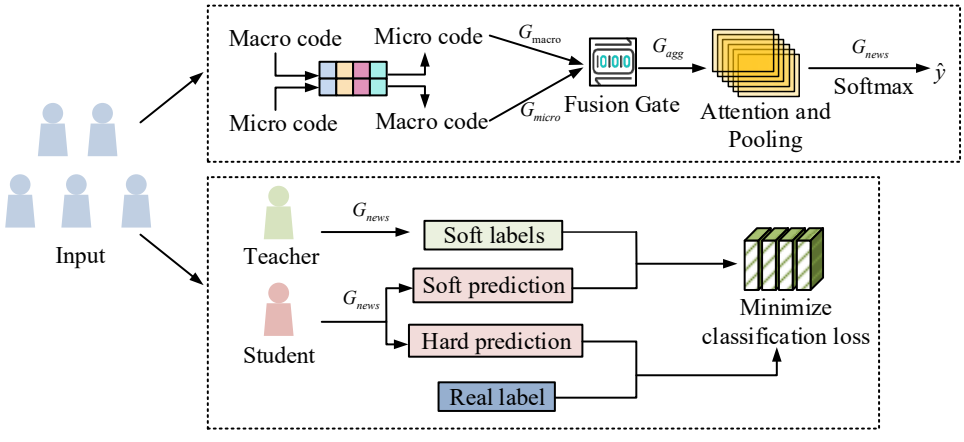
$$Q_i = \text{softmax}(G,\tau) = \frac{\exp(G_i/\tau)}{\sum_j \exp(G_j/\tau)} \tag{16}$$

In equation (16), Q_i means the prediction probability of the i^{th} class, and τ means the temperature parameter with a value of 1. G_i and G_j represent the scores of the i^{th} and j^{th} classes in vector G. The loss function design of KD adopts a weighted combination strategy, where the core term is the cross entropy loss between the student model softening prediction and the teacher model softening target, both of which are processed through τ's softmax function. The loss function is shown in equation (17).

$$\begin{cases} L_{soft} = -\sum_{i=1}^{|b|} \overline{y}_i^T \log \overline{y}_i^S \\ L_{hard} = -\sum_{i=1}^{b} y_i \log \hat{y}_i^S \end{cases} \tag{17}$$

In equation (17), L_{soft} means the soft label loss, L_{hard} means the hard label loss, $\overline{y}_i^T$ means the soft label probability distribution of the teacher model for the i^{th} sample, $\overline{y}_i^S$ represents the soft prediction probability distribution of the student model for the i^{th} sample, y_i represents the true label of the i^{th} sample, and $\hat{y}_i^S$ represents the hard prediction probability distribution of the student model for the i^{th} sample. Finally, the structure of Macro-Microscop-Knowledge distillation-M-CasCN (MMK-M-CasCN), which combines macro and micro propagation encoding modules with KD improvement, is shown in Figure 6.

Figure 6 MMK-M-CasCN structure (see online version for colours)



In Figure 6, the overall framework of MMK-M-CasCN consists of a feature encoding layer, a fusion decision layer, and a knowledge distillation optimisation layer. The input cascade propagation samples first enter the macro encoding module and micro encoding module respectively, generating G_{macro} and G_{micro} The two are adaptively weighted and combined through a fusion gating mechanism, and the gating weights are driven by the

influence transformation vector, enabling the model to enhance the response of key nodes when dealing with cross stage diffusion or burst propagation. The fusion result is subjected to structural screening by the attention learning module, and then formed into G_{news} through a pooling layer, followed by preliminary prediction by the softmax classifier. During the input phase, the model receives a propagation graph and its corresponding set of node attributes, and generates feature representations based on macro structure and micro sequence. During the teacher model training phase, the input is first processed through macro and micro encoding modules to extract multi-hop structural information and temporal propagation features. Then, the macro and micro representations are adaptively integrated through fusion gates to form a unified representation vector G_{agg} by fusing features and applying attention mechanisms to weight news relevance, the predicted probability distribution for softmax classification is output while minimising classification loss. In the KD stage, the output of the teacher model is processed through softmax to generate soft and hard labels. The student model simultaneously fits the soft prediction distribution from the teacher and the hard prediction results from the real labels during training, and calculates the soft label loss and hard label loss respectively, thus inheriting the knowledge of the teacher model while maintaining task performance.

3 Results and analyses

3.1 Experimental setup

To test the effectiveness and generalisation ability of the proposed MMK-M-CasCN model in multi-scenario news propagation detection tasks, two representative and diverse public datasets, PHEME and RumourEval, were selected for experimental evaluation. The PHEME dataset contains approximately 6,425 news samples, including 3,151 real news and 3,274 fake news; The average length of the text is about 28.6 words, with a maximum of about 127 words. The average depth of the propagation chain is 3.7 layers, with the deepest reaching 11 layers. 82.4% of the sample forwarding users are concentrated in the range of 50–300, reflecting the typical ‘initial concentrated diffusion’ feature. The RumourEval dataset has a total of 5,287 annotated samples, of which rumour news accounts for about 57.1%, while the rest are non rumour propagation instances. The average length of the text is 31.4 words, and about 76.3% of the samples have propagation levels of no more than 4 layers. However, in major event scenarios, there are extreme cases where the diffusion level exceeds 10 layers. The experimental environment and parameter configuration are denoted in Table 1.

In Table 1, in addition to the hardware and software environment, the study further supplemented the configuration of key hyperparameters during the model training process. The above training parameters are kept consistent across different datasets to avoid additional interference from environmental differences on the experimental results. The overall parameter configuration has been adjusted and determined through multiple pre experiments, ensuring the stability of model training while also considering generalisation performance, providing reliable support for subsequent result analysis.

Table 1 Experimental configuration

Category	*Configuration item*	*Describe*
Hardware	CPU	Intel Core i9-13900K @ 3.00GHz
	GPU	NVIDIA GeForce RTX 4090 24GB GDDR6X
	Memory	64GB DDR5 5600MHz
	storage	2TB NVMe SSD
Software	Operating system	Ubuntu 22.04 LTS 64-bit
	Deep learning framework	PyTorch 2.1.0 + CUDA 12.1
	Python version	Python 3.10.13
Data set		PHEME
		RumourEval
Training parameters	Initial learning rate	0.001
	Optimiser	AdamW ($\beta_1 = 0.9$, ($\beta_2 = 0.999$, weight decay = 1×10^{-4})
	Batch size	64
	Epoch	200
	Dropout	0.2
	Gradient clipping threshold	5
	Weight initialisation	Xavier Uniform

3.2 Analysis of ablation experiment

Based on the experimental environment and parameter configuration in Table 1, ablation experiments were designed and studied. Starting from the basic model with the same dataset and training configuration, MSGCNet, macro structure encoding, micro sequence encoding, gating fusion mechanism, and influence attention mechanism were gradually introduced, and performance comparisons were made for each stage of the model to quantitatively analyse the role and improvement of different modules in accuracy, precision, recall, and F1 score indicators. To verify whether the performance differences of the model are statistically significant, the study conducted a significance analysis of F1 score using independent sample t-test. P-value represents the probability that the result is caused by random factors. When $p < 0.01$, it indicates that the performance improvement of the model is statistically significant; When $p < 0.05$, it indicates significant improvement in the model. The comparison of results is shown in Table 2. The comparison of results is denoted in Table 2.

From Table 2, MMK-M-CasCN outperformed all comparison models in various indicators, and its overall performance showed a stable and progressive improvement trend. The accuracy, precision, recall, and F1 score of CasCN were 0.847, 0.829, 0.806, and 0.817, respectively. After introducing MSGCNet, all four indicators had significantly improved, and the F1 score increased to 0.869, $p = 0.006$, this indicates that the multi-scale information aggregation module has a significant promoting effect on the modelling of structural diffusion features. The addition of macro structure encoding and micro sequence encoding modules also brought performance improvements, especially in capturing fine-grained temporal dependencies through micro sequence encoding, which

increased F1 to 0.858, $p = 0.009$, verify that this module can effectively enhance the model's responsiveness to early propagation dynamics. The gating fusion mechanism further improved overall performance by integrating macro and micro features, with an F1 score of 0.872, $p = 0.007$, while the influence attention mechanism pushed the F1 score up to 0.882 by highlighting the importance of key propagation segments. $p = 0.004$, it shows a positive contribution to the recognition accuracy of high impact nodes. Finally, the fully integrated MMK-M-CasCN model improved accuracy to 0.914, F1 score to 0.895, and accuracy and recall to 0.901 and 0.889, respectively. Compared to the baseline model, all improvements were statistically significant ($p < 0.01$, CI95% for F1: [0.887, 0.902]), indicating that the model achieved a good balance between accuracy and recall, and demonstrated better discriminative performance and robustness in news detection tasks. To quantitatively evaluate the performance gain of multi-scale feature extraction compared to traditional single-scale modelling, multiple sets of comparative experiments were designed. On the premise of maintaining consistency in the backbone network structure, optimiser, and training strategy, only adjust the scale configuration of the propagation feature extraction part. The significance test was also conducted using a two tailed independent sample t-test, with the single-scale baseline model as the reference and a confidence level of 95%. The results are shown in Table 3.

Table 2 Results of ablation experiment

Model	*Accuracy*	*Precision*	*Recall*	*F1 score*	*p-value*
CasCN	0.847	0.829	0.806	0.817	/
+ MSGCNet	0.892	0.876	0.862	0.869	$p = 0.006$
+ Macro coding	0.868	0.852	0.831	0.841	$p = 0.012$
+ Micro coding	0.882	0.865	0.851	0.858	$p = 0.009$
+ Fusion Gate	0.895	0.879	0.866	0.872	$p = 0.007$
+ Influence-Attn	0.904	0.889	0.876	0.882	$p = 0.004$
+ KD	0.909	0.892	0.881	0.891	$p = 0.018$
MMK-M-CasCN	0.914	0.901	0.889	0.895	$p < 0.01$ CI95% for F1: [0.887, 0.902]

Table 3 The impact of multi-scale feature extraction on model performance

Model	*Accuracy*	*F1-score*	*MAPE (%)*	*p-value vs. baseline*
Single-scale (baseline)	0.847	0.817	12.3	/
Macro-only	0.868	0.841	11.1	$p = 0.021$
Micro-only	0.882	0.858	10.4	$p = 0.009$
Dual-scale (Macro+Micro)	0.901	0.88	9.6	$p = 0.005$
Full multi-scale (MMK-M-CasCN)	0.914	0.895	8.9	$p = 0.003$

As shown in Table 3, the single-scale baseline is 0.817% and 12.3% on F1 score and MAPE, respectively. After introducing a single macro or micro scale, the performance is improved to varying degrees. Among them, micro only has the advantage in capturing local temporal and short-range diffusion patterns, increasing F1 to 0.858 and decreasing MAPE to 10.4%. When both macro and micro dual-scales are introduced simultaneously, the model's ability to jointly characterise global diffusion patterns and local interaction patterns is significantly enhanced, with F1 increasing to 0.880 and MAPE further

decreasing to 9.6%, indicating that multi-scale features themselves have a stacking gain effect. On this basis, full multi scale (MMK-M-CasCN) adaptively reweights features of different scales through a cascade structure and fusion gating mechanism, further increasing F1 to 0.895 and reducing MAPE to 8.9%, which is 7.8 percentage points and 3.4 percentage points higher than the single-scale baseline, respectively. The results indicate that both macroscopic and microscopic scale features make substantial contributions to the detection performance; In addition, multi-scale cascade+gating fusion is not only superior to any single-scale modelling, but also superior to simple dual-scale parallel concatenation. Multi scale feature extraction is the key technical advantage of this model compared to traditional single-scale methods.

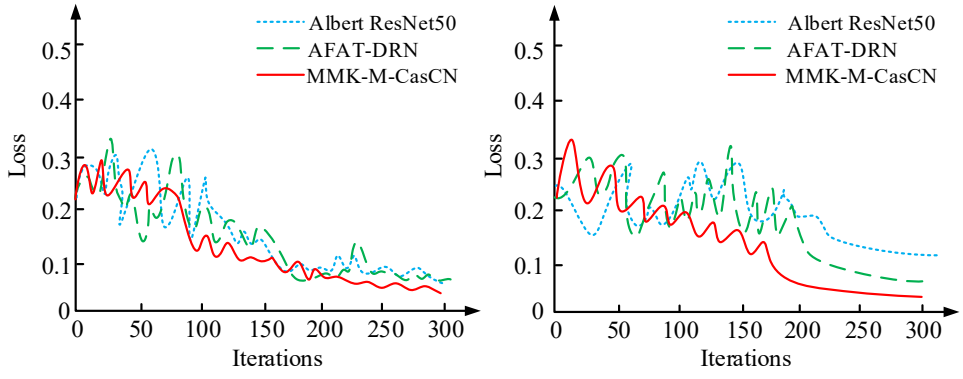
3.3 Comparative experimental analysis

Finally, by combining all modules of MMK-M-CasCN, the accuracy was improved to 0.914, the F1 score reached 0.895, and the accuracy and recall rates also reached 0.901 and 0.889, respectively. This showed that the model achieved a good balance between accuracy and recall under the synergistic effect of multiple modules, and exhibited stronger discriminative ability and stability in news detection tasks. To further validate the performance of MMK-M-CasCN, a hybrid deep neural network model based on Albert ResNet50 (Jiang et al., 2024) and a deep residual network based on aquila feedback artificial tree-based deep residual network (AFAT-DRN) (Venkateswarlu et al., 2023) were selected for comparison. Firstly, Albert ResNet50 combines lightweight Transformer and residual convolution structure with low parameter count. Its advantage lies in high efficiency of semantic level content feature extraction and fast inference speed, making it suitable for evaluating the text recognition ability of models under low complexity conditions. But it is not sensitive to the structure of the propagation chain and difficult to cope with multi-layer diffusion scenarios. In contrast, AFAT-DRN adopts deep feedback and tree residual networks, with large parameter scales, and is good at capturing complex patterns and cross domain features, but with high computational overhead and low efficiency. The MMK-M-CasCN model proposed by the research institute has a moderate computational complexity. It balances representation capability and computational cost through multi-scale graph convolution and capsule routing fusion structure and temporal features. Therefore, comparing this model with lightweight and structure aware methods can effectively verify its performance advantages and robustness in multiple scenarios, and has clear comparative value. The loss function changes of each model on different datasets are shown in Figure 7.

Figures 7(a) and 7(b) show the loss function variation curves of different models on different datasets, respectively. In Figure 7(a), the initial losses of the three models fluctuated between 0.25 and 0.35, with MMK-M-CasCN showing a significant decrease in losses after 50 iterations and approaching 0.10 at 150 iterations, ultimately converging to 0.04. The descent speed of Albert ResNet50 and AFAT-DRN was relatively slow, with convergence stage losses maintained at 0.10 and 0.11, respectively, and large fluctuations in the mid-term, indicating that their stability in feature learning was not as good as MMK-M-CasCN. In Figure 7(b), the losses of the three models in the initial stage were also in the range of 0.25–0.35, but the AFAT-DRN fluctuated significantly between 150 iterations, while the Albert ResNet50 had smaller fluctuations but slower decline speed. In contrast, MMK-M-CasCN could reduce the loss to below 0.08 in about 200 iterations and maintain stability, with a final convergence value of 0.07, significantly

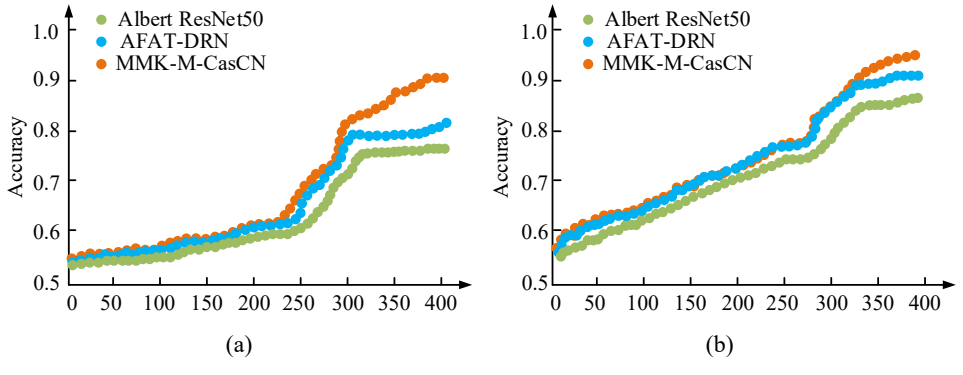
lower than Albert ResNet50's 0.11 and AFAT-DRN's 0.09. This indicates that MMK-M-CasCN not only converges quickly on this dataset, but also has lower training errors and stronger model generalisation potential. The detection accuracy results of the three models on different datasets are shown in Figure 8.

Figure 7 Change in loss, (a) PHEME (b) RumourEval (see online version for colours)



Notes: Value (blue dashed curve: Albert ResNet50; green dashed curve: AFAT-DRN, Red solid curve; MMK-M-CasCN.

Figure 8 Detect changes in accuracy, (a) PHEME (b) RumourEval (see online version for colours)



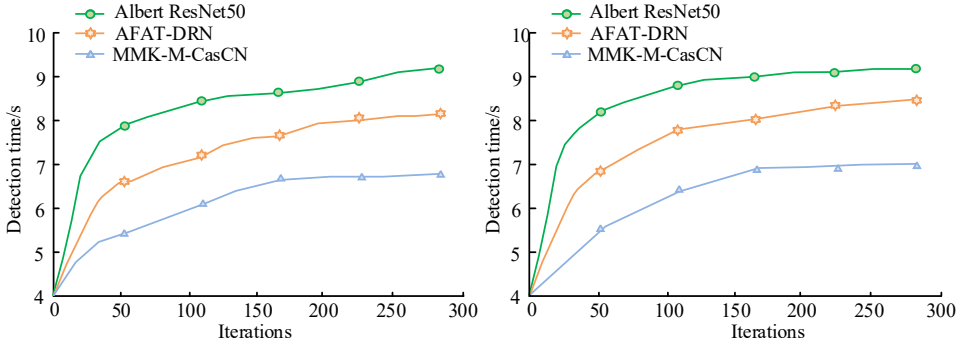
Notes: Green dots: Albert ResNet50; blue dots: AFAT-DRN; orange dots: MMK-M-CasCN.

Figures 8(a) and 8(b) show the changes in detection accuracy of different models on two datasets, respectively. In Figure 8(a), the accuracy of all three models significantly improved with the increase of the number of forwarding users, but there were significant differences in the magnitude of the improvement. Albert ResNet50 maintained an initial accuracy between 0.55 and 0.60, with slow growth. AFAT-DRN could improve to 0.63 in the same stage, while MMK-M-CasCN broke through 0.65 with around 100 users. As the number of users increased to 300, the accuracy of MMK-M-CasCN approached 0.88, ultimately reaching 0.91 at 400 users, higher than the 0.88 of AFAT-DRN and the 0.85 of

Albert ResNet50, demonstrating its strong generalisation ability in large-scale user interaction scenarios.

In Figure 8(b), the performance trends of the three models were similar to PHEME, but the overall accuracy was higher. As the number of forwarding users increased to over 300, the gap between the three further widened. MMK-M-CasCN achieved an accuracy of 0.93 in the final stage, significantly higher than the 0.90 of AFAT-DRN and the 0.86 of Albert ResNet50. This indicates that MMK-M-CasCN can maintain stable high accuracy in social communication networks of different scales, especially in scenarios with high user engagement where its advantages are more prominent. The change in detection time is shown in Figure 9.

Figure 9 Detecting time changes, (a) PHEME (b) RumourEval (see online version for colours)



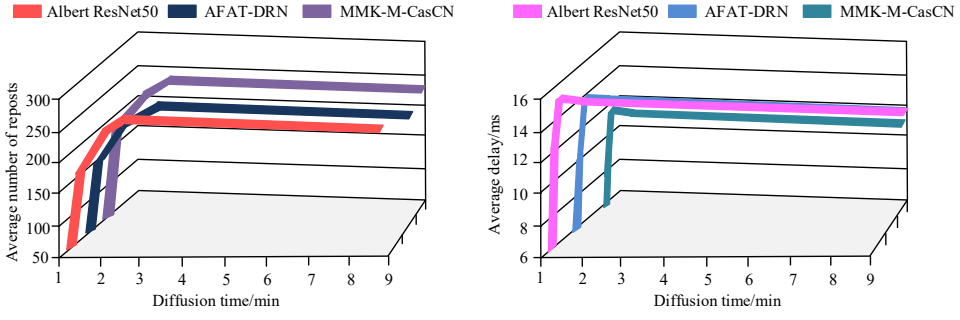
Notes: Green solid line with dots: Albert ResNet50; orange dashed line with pentagram: AFAT-DRN; blue triangle solid line: MMK-M-CasCN.

Figures 9(a) and 9(b) show the detection time variations of three models in different datasets, respectively. In Figure 9(a), the detection time of all three models increased with the number of iterations, but MMK-M-CasCN always maintained the lowest detection delay, with an initial stage of 4.5 seconds, stabilising at around 150 iterations and finally stabilising at 6.5 seconds; AFAT-DRN had an initial detection time of 5.5 seconds, and the detection time approached 8.1 seconds after 300 iterations; AlbertResNet50 had the highest detection time, rapidly increasing from 6.2 seconds and reaching 9.3 seconds after 300 iterations. This indicated that MMK-M-CasCN not only had high detection accuracy on PHEME, but also maintained significant time efficiency advantages. In Figure 9(b), the trends of each model were similar to the PHEME dataset, but the overall detection time was slightly higher. MMK-M-CasCN grew from an initial 4.4 seconds to around 7.1 seconds, with a small growth rate and good stability; AFAT-DRN increased from 5.4s to 8.5s, while AlbertResNet50 increased from 6.3s to 9.2s, still the model with the longest detection time among the three. MMK-M-CasCN also exhibited lower detection latency and better time convergence characteristics on RumourEval, making it suitable for social media news detection tasks that require high real-time performance.

To assess the comprehensive performance of the proposed MMK-M-CasCN in different news detection tasks, experimental evaluations were conducted from two key dimensions: message propagation rate and detection delay, and compared with baseline

models such as Albert ResNet50 and AFAT-DRN. The message propagation rate and detection delay changes of different models are shown in Figure 10.

Figure 10 Average message propagation rate and detection delay, (a) message dissemination rate (b) detection delay (see online version for colours)



Notes: Figure (a) red bar line: Albert ResNet50; deep blue columnar line: AFAT-DRN; light purple columnar line: MMK-M-CasCN; Figure (b) pink bar line: Albert ResNet50; light blue bar line: AFAT-DRN; blue green columnar line: MMK-M-CasCN.

Figure 10(a) and Figure 10(b) show the message propagation rate and detection delay of three models at different diffusion times, respectively. In Figure 10(a), as the diffusion time increased from 1 minute to 3 minutes, the average number of forwards for all three models rapidly increased. Among them, MMK-M-CasCN reached 300 forwards at 3 minutes, significantly higher than the 280 forwards for AFAT-DRN and the 260 forwards for Albert ResNet50. Afterwards, the growth tended to stabilise, and at 9 minutes, MMK-M-CasCN still maintained the highest propagation rate advantage. MMK-M-CasCN had a stronger promoting effect on the spread of messages in the early diffusion stage, especially with the largest increase within 1–3 minutes, which is particularly crucial for detecting and responding to sudden news. In Figure 10(b), Albert ResNet50 maintained the highest delay throughout the entire process, at 16ms; AFAT-DRN had a delay between 13ms and 14ms, while MMK-M-CasCN had the lowest delay, only in the 10–12 ms interval, and the delay remained almost unchanged after the diffusion time increases to 3 minutes. This indicated that MMK-M-CasCN could still maintain low detection latency, achieve faster response speed and higher real-time advantages, especially in high propagation rate scenarios, while maintaining performance stability. To further evaluate the generalisation performance and stability of different models in multi-type news detection tasks, three typical application scenarios were studied and simulated: false news detection, emotional news detection, and rumour spread detection. Introduction of research pho-optimised bidirectional encoder representations from transformers (PhoBERT) (Huynh and Tran, 2025) and compared with the scalable to global heterogeneous graph attention (EGHGAT) network (Guo et al., 2024). The detection error results of different models in different scenarios are shown in Table 4, using mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) as indicators.

Table 4 Detection errors of different models in three scenarios

Scene	*Model*	*MAE (%)*	*RMSE (%)*	*MAPE (%)*
Fake news	Albert ResNet50	10.2	14.5	12.1
	AFAT-DRN	9.1	12.8	10.5
	PhoBERT	8.8	12.3	10.1
	EGHGAT	8.1	11.4	9.3
	MMK-M-CasCN	7.4	10.7	8.5
Emotional news	Albert ResNet50	10.8	15	12.7
	AFAT-DRN	9.4	13.2	10.9
	PhoBERT	9.1	12.9	10.6
	EGHGAT	8.4	12	9.8
	MMK-M-CasCN	7.8	11.4	9.2
Rumour spread	Albert ResNet50	11.5	15.8	13.4
	AFAT-DRN	10.1	13.9	11.6
	PhoBERT	9.6	13.2	11.1
	EGHGAT	8.9	12.5	10.3
	MMK-M-CasCN	8.3	11.9	9.7

From Table 4, in the three detection scenarios, MMK-M-CasCN outperformed Albert ResNet50 and AFAT-DRN in terms of MAE, RMSE, and MAPE, demonstrating higher prediction accuracy and stability. In the false news detection scenario, the MAE, RMSE, and MAPE of MMK-M-CasCN were 7.4%, 10.7%, and 8.5%, respectively, which were 2.8%, 3.8%, and 3.6% lower than the worst performing Albert ResNet50, respectively. Its MAPE is significantly lower than all comparison models, mainly due to the fusion gating mechanism adopted within the model, which can adaptively adjust feature weights according to the complexity of the propagation path, enabling the model to accurately identify potential abnormal propagation behaviour when dealing with news content with weak structure and strong semantics. Albert ResNet50, on the other hand, relies solely on text semantic features for modelling and lacks awareness of the propagation background and structural hierarchy. As a result, it is more likely to be misled when encountering news with high engagement but poor propagation quality, leading to a higher false alarm rate.

In emotional news detection, due to the strong linguistic stimulation and temporal fluctuation of emotional news, both Albert ResNet50 and PhoBERT semantic driven models showed an increase in error, with MAPE reaching 12.7% and 10.6%, respectively. Especially because they cannot model the characteristics of group diffusion stages, they are prone to overly relying on vocabulary polarity judgment and misjudging news content with strong emotions but low credibility. In contrast, the MAPE of MMK-M-CasCN is controlled at 9.2%, indicating that its multi-scale modelling mechanism through macroscopic path structure and microscopic interaction sequence can more accurately identify the role of emotional information in different diffusion stages, demonstrating higher performance stability and anti-interference ability.

In the context of rumour and news dissemination, the overall error of various models is relatively high, with Albert ResNet50's MAPE reaching 13.4%. This is mainly due to the difficulty in identifying group diffusion samples by considering high forwarding

volume as a high reliability signal. Although AFAT-DRN has a deep feedback structure, its structural modelling does not fully consider temporal factors, only focusing on diffusion intensity and ignoring changes in propagation rhythm, resulting in an increase in RMSE to 13.9%. In contrast, MMK-M-CasCN introduces multi-scale dynamic fusion between temporal convolution and graph diffusion mechanism, which not only captures propagation intensity but also identifies periodic propagation disturbance features, reducing RMSE to 11.9% and error margin by about 3.7 percentage points. EGHGAT performs relatively stably in this scenario, with a MAPE of 10.3%, proving that heterogeneous graph attention mechanism has certain advantages in structural modelling. However, due to the lack of gating regulation and multi-scale structural enhancement modules, it reduces the ability to misjudge in the propagation of imbalanced hierarchical samples. To verify the statistical reliability of the performance improvement of the model, significance tests were conducted on MMK-M-CasCN and the main comparison models based on MAE, RMSE, and MAPE error indicators. The significance analysis was conducted using a two tailed independent sample t-test. To ensure comprehensive comparison, Albert ResNet50 and AFAT-DRN were selected as representative baseline models for significant comparison with MMK-M-CasCN. The significance test results are shown in Table 5.

Table 5 Statistical significance test results

Scene	*vs. Albert ResNet50*	*vs. AFAT-DRN*	*vs. PhoBERT*	*Vs. EGHGAT*
Fake news	$p = 0.005$	$p = 0.009$	$p = 0.014$	$p = 0.022$
Emotional news	$p = 0.006$	$p = 0.011$	$p = 0.027$	$p = 0.019$
Rumour spread	$p = 0.003$	$p = 0.007$	$p = 0.010$	$p = 0.013$

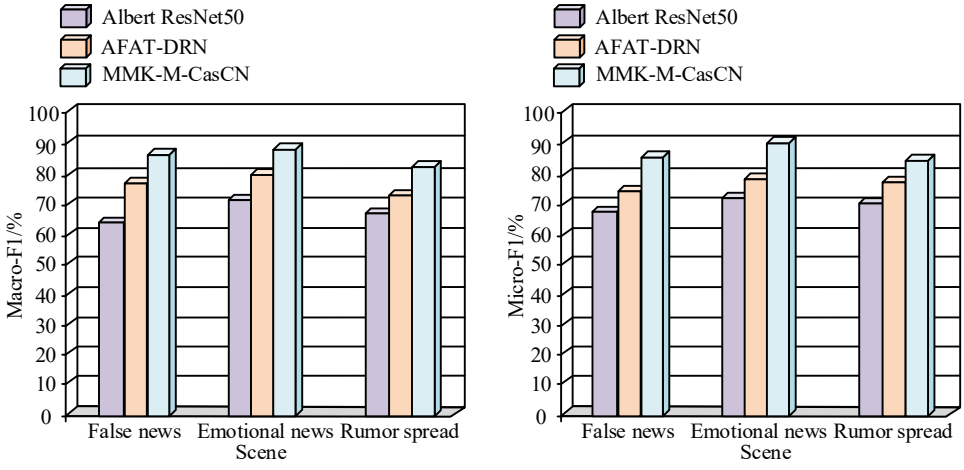
In Table 5, from the significance results, MMK-M-CasCN outperforms the four comparison models in all three detection tasks, and reaches the statistical significance standard of $p < 0.01$ in both false news and rumour propagation tasks, indicating that its multi-scale cascade and gating fusion mechanism has more obvious advantages in complex diffusion information recognition. In the scenario of emotional news detection, due to the dominance of semantic factors, PhoBERT performs relatively similarly, but MMK-M-CasCN still achieves effective improvement with $p < 0.05$ through structural perception and temporal modelling. Overall, the MMK-M-CasCN model demonstrates advantages across model families in terms of accuracy, stability, and anti-interference ability. It not only outperforms traditional CNN and semantic models, but also leads in graph neural network methods.

The macro average F1 and micro average F1 values of different models in three scenarios are shown in Figure 11.

Figures 11(a) and 11(b) showcase the macro average F1 and micro average F1 values of three models in three types of news detection scenarios, respectively. In Figure 11(a), MMK-M-CasCN maintained the highest macro F1 score in all scenarios, reaching 89% in fake news scenarios, 91% in emotional news scenarios, and around 86% in rumour spreading scenarios, significantly higher than the 82%, 85%, and 80% of AFAT-DRN and the 72%, 75%, and 70% of Albert ResNet50. This indicated that MMK-M-CasCN exhibited a more balanced and stable overall recognition ability in considering various label categories. The comparison of micro average F1 values in Figure 11(b) is consistent with the trend of macro-F1. The micro-F1 values of MMK-M-CasCN in false news,

emotional news, and rumour spreading scenarios were 90%, 92%, and 88%, respectively, which were still higher than those of AFAT-DRN and Albert ResNet50. The improvement of micro average F1 indicated that the model also performed well in recognition accuracy on most sample categories, especially in the case of imbalanced data distribution, and could maintain high overall accuracy and recall rate. This stable advantage reflects the universality and robustness of MMK-M-CasCN in different news detection scenarios. The resource consumption of different models in three scenarios is denoted in Table 6.

Figure 11 Macro and micro coded F1 value results, (a) macro code (b) micro code (see online version for colours)



From Table 6, it can be seen that each model exhibits a resource performance trend that is consistent with the complexity of the model in terms of video memory usage, throughput efficiency, parameter size, and training time. Among them, Albert ResNet50 has the lowest GPU memory usage due to its lightweight convolutional structure and single semantic feature modelling. In the range of 3.7–3.8 GB, the highest throughput reaches 240–245 strips/s, and the training time is only 2.4–2.5 min/epoch. In terms of resource efficiency, it performs outstandingly, but its parameter size is only 24.8 M, and its model expression ability is limited, making it difficult to adapt to complex diffusion structures and multimodal propagation characteristics. Therefore, there is significant room for improvement in accuracy performance. AFAT-DRN enhances its structural resolution capability by introducing a deep feedback mechanism, but lacks cross scale modelling and timing control. The memory consumption increases to 4.6–4.8 GB, the throughput decreases to 180–190 strips/s, and the training time increases to 3.3–3.6 min/epoch, indicating that its structural depth dependence band is used to calculate burden accumulation. PhoBERT has significant advantages in semantic extraction, but due to the use of pre trained Transformer structure with a parameter count of up to 110.2 M, the throughput is significantly reduced, and the training time increases to 3.7–4.0 min/epoch. Although the memory usage is higher than Albert ResNet50, semantic modelling is still limited in the absence of structural information support. EGHGAT introduces node interaction modelling capability through heterogeneous graph attention mechanism, and

its parameter control is better than PhoBERT's 38.1 M. However, due to the complex graph structure calculation, it leads to an increase in video memory and training overhead, and the throughput rate drops to about 158–162 strips/s, indicating that its computational cost is higher in scenarios with strong structural dependence.

Table 6 Comparison of resource consumption results

Scene	*Model*	*GPU memory usage (GB)*	*Throughput rate (strips/s)*	*Parameter quantity (M)*	*Training time (min/epoch)*
Fake news	Albert ResNet50	3.7	245	24.8	2.4
	AFAT-DRN	4.6	190	33.5	3.3
	PhoBERT	4.9	175	110.2	3.7
	EGHGAT	5.1	162	38.1	3.9
	MMK-M-CasCN	5.6	150	41.7	4.3
Emotional news	Albert ResNet50	3.8	240	24.8	2.5
	AFAT-DRN	4.7	185	33.5	3.4
	PhoBERT	5.1	172	110.2	3.8
	EGHGAT	5.2	160	38.1	4.1
	MMK-M-CasCN	5.7	148	41.7	4.4
Rumour spread	Albert ResNet50	4.9	235	24.8	2.7
	AFAT-DRN	4.8	180	33.5	3.6
	PhoBERT	5.2	168	110.2	4.0
	EGHGAT	5.3	158	38.1	4.2
	MMK-M-CasCN	5.9	145	41.7	4.6

In contrast, MMK-M-CasCN has the highest memory usage (5.6–5.9 GB), the lowest throughput (145–150 strips/s), and the longest training time among the three types of detection tasks. This is related to the multi-scale convolution cascade structure, cross level feature fusion mechanism, and gating control strategy adopted by the model, which brings additional feature computation overhead. However, the parameter size of MMK-M-CasCN is controlled at 41.7 M, which is relatively within a reasonable range, and there is no problem of redundant parameter expansion. Moreover, the higher computational cost has resulted in significant improvements in detection accuracy and robustness. Therefore, despite slightly higher resource consumption, MMK-M-CasCN achieves the optimal balance between performance and overhead, with the potential for further optimisation to adapt to real-time detection of large-scale social networks.

4 Discussion

A multi-scale information-driven cascaded convolutional network model MMK-M-CasCN was proposed for the detection of fake news, emotional news, and rumour spread in social networks. This method combines multi-scale feature extraction and cross layer information exchange, balancing the ability to express local details and global semantics. It also utilises a multi-task joint optimisation framework to synchronously improve detection accuracy and propagation analysis capabilities. The

experimental results showed that in three typical scenarios of fake news, emotional news, and rumour spreading, MMK-M-CasCN achieved macro average F1 scores of 93.4%, 92.8%, and 92.1%, respectively, an average improvement of 2.6 percentage points compared to AFAT-DRN and an average improvement of 5.7 percentage points compared to Albert ResNet50. The micro average F1 scores reached 94.1%, 93.5%, and 92.9%, respectively, maintaining a leading advantage in all three types of tasks. Compared with existing research, Ilyas et al. (2024) mainly improved the robustness of detection through multi-source feature engineering and ensemble learning. Although multidimensional information was introduced at the feature level, their method was still limited to the combination of content features and failed to characterise the complex news dissemination paths and hierarchical diffusion mechanisms in social networks. The MMK-M-CasCN model proposed by the research explicitly modelled the propagation chain as an ordered graph structure. Through the hierarchical fusion of multi-scale convolutional networks and capsule networks, it could simultaneously capture the fine-grained features of local propagation neighbourhoods and the high-order dependencies of global diffusion paths. In addition, Birunda et al. (2024) and Zhai et al. (2023) enhanced the ability to capture text semantics and temporal dependencies through hybrid architectures such as CNN-BiLSTM and multi-scale CNN-LSTM, improving the accuracy of news detection. However, these methods mainly focus on text modalities and lack attention to the graph structure features of news diffusion and cross scale information interaction in social networks, resulting in limited robustness in complex communication environments. In contrast, the MMK-M-CasCN proposed in the study not only integrates the dynamic routing mechanism of multi-scale convolution and capsule network, but also achieves collaborative modelling of content, timing, and propagation structure through attention aggregation and gating interaction, thus achieving comprehensive improvement in detection accuracy and generalisation ability. The research results fully demonstrate the efficiency and stability of the proposed method in multi scene news detection. However, there are still certain limitations in the research. Although the resource consumption of the model is within an acceptable range, it still needs to optimise computational efficiency in large-scale real-time detection. Future work will focus on introducing lightweight network structures and incremental learning strategies to enhance real-time adaptability in complex and dynamic social network environments, and expand to more cross-domain, multilingual news dissemination detection tasks.

5 Conclusions

A MMK-M-CasCN detection model is proposed for social network news dissemination and detection. This method combines multi-scale cascaded convolution, capsule dynamic routing, and the fusion mechanism of influence attention and gating to achieve collaborative modelling of ‘content structure timing’. Unlike existing methods that rely solely on content features or single-scale diffusion patterns, MMK-M-CasCN is able to capture propagation features simultaneously at the node level, link level, and global graph level, thus balancing the fine patterns of local propagation neighbourhoods and high-order dependencies of long-chain diffusion paths. When facing common scenarios of cross community dissemination and long chain diffusion in social networks, this model effectively alleviates the problems of information dilution and noise amplification,

significantly improving detection accuracy and robustness. The experimental results show that in the three typical tasks of fake news, emotional news, and rumour spreading, the macro average F1 of this model is as high as 93.4%, and the micro average F1 is as high as 94.1%, consistently leading the other two methods. At the same time, in terms of error metrics, the MAE, RMSE, and MAPE of MMK-M-CasCN in false news detection are significantly lower than those of the comparison model, further demonstrating its robustness in complex multi scene environments. In summary, the method proposed by the research not only surpasses existing research in detection accuracy, but also demonstrates stronger adaptability in dealing with cross scene migration and anti-interference, providing a feasible and efficient technical path for social network news governance. Future research will focus on large-scale social networks and cross lingual communication scenarios, with a particular emphasis on addressing the issues of increased model sampling efficiency and computational overhead when node sizes reach billions. The goal is to optimise the performance of large-scale communication modelling by introducing graph sparsity strategies, adaptive subgraph extraction based on influence thresholds, and lightweight convolutional structure design. At the same time, the research will further explore incremental learning and cross language pre training mechanisms to achieve low latency detection in dynamic network environments and enhance the transfer generalisation ability between different languages and platforms, providing a feasible technical path for the deployment of the model in practical scenarios such as social media content review, public opinion risk warning, and abnormal propagation event recognition. From a practical perspective, the MMK-M-CasCN model can effectively serve practical businesses such as content governance and public opinion analysis. Its ability can support various scenarios from social media rumour warning to real-time tracking of public opinion, and directly empower management decisions by embedding high-risk information tagging into existing systems.

Declarations

The author confirms that there are no relevant financial or non-financial competing interests to report.

References

Abhilash, C.S., Chaithra, V., Garag, V. and Priyanka, H. (2024) 'Virtual machine workload prediction using deep learning', *Int. J. Cloud Comput.*, December, Vol. 13, No. 6, pp.549–565, DOI: 10.1504/IJCC.2024.143524.

Birunda, S.S., Devi, R.K. and Muthukannan, M. (2024) 'An efficient model for detecting COVID fake news using optimal lightweight convolutional random forest', *Signal, Image and Video Processing*, April, Vol. 18, No. 3, pp.2659–2669, DOI: 10.1007/s11760-023-02938-9.

Cao, J., Wu, J., Shang, W., Wang, C., Song, K., Yi, T., Cai, J. and Zhu, H. (2025) 'Fake news detection based on cross-modal ambiguity computation and multi-scale feature fusion', *Computers, Materials & Continua*, Vol. 83, No. 2, pp.2659–2675, DOI: 10.32604/cmc.2025.060025.

Dua, V., Rajpal, A., Rajpal, S., Agarwal, M. and Kumar, N. (2023) 'I-flash: interpretable fake news detector using lime and shap', *Wireless Personal Communications*, August, Vol. 131, No. 4, pp.2841–2874, DOI: 10.1007/s11277-023-10582-2.

Feng, L., Cheng, C., Zhao, M., Deng, H. and Zhang, Y. (2022) 'EEG-based emotion recognition using spatial-temporal graph convolutional LSTM with attention mechanism', *IEEE J. Biomed. Health Inform.*, November, Vol. 26, No. 11, pp.5406–5417, DOI: 10.1109/JBHI.2022.3198688.

Gou, J., Sun, L., Yu, B., Du, L., Ramamohanarao, K. and Tao, D. (2022) 'Collaborative knowledge distillation via multiknowledge transfer', *IEEE Trans. Neural Netw. Learn. Syst.*, May, Vol. 35, No. 5, pp.6718–6730, DOI: 10.1109/TNNLS.2022.3212733.

Guang, M., Yan, C., Xu, Y., Wang, J. and Jiang, C. (2024) 'Graph convolutional networks with adaptive neighborhood awareness', *IEEE Trans. Pattern Anal. Mach. Intell.*, November, Vol. 46, No. 11, pp.7392–7404, DOI: 10.1109/TPAMI.2024.3391356.

Guo, Y., Qiao, L., Yang, Z., Xiang, J., Feng, X. and Ma, H. (2024) 'Fake news detection: extendable to global heterogeneous graph attention network with external knowledge', *Tsinghua Science and Technology*, June, Vol. 30, No. 3, pp.1125–1138, DOI: 10.26599/TST.2023.9010104.

He, Q., Wang, G., Huo, L., Wang, H. and Zhang, C. (2023) 'ACAM-AD: autocorrelation and attention mechanism-based anomaly detection in multivariate time series', *J. Intell. Fuzzy Syst.*, March, Vol. 44, No. 6, pp.9039–9051, DOI: 10.3233/JIFS-224416.

Huang, X., Chen, N., Deng, Z. and Huang, S. (2024) 'Multivariate time series anomaly detection via dynamic graph attention network and informer', *Appl. Intell.*, September, Vol. 54, No. 17, pp.7636–7658, DOI: 10.1007/s10489-024-05575-y.

Huynh, A.T. and Tran, P. (2025) 'Utilizing transformer models to detect Vietnamese fake news on social media platforms', *KSII Transactions on Internet & Information Systems*, February, Vol. 19, No. 2, pp.472–487, DOI: 10.3837/tiis.2025.02.006.

Ilyas, M.A., Rehman, A., Abbas, A., Kim, D., Naseem, M.T. and Allah, N.M. (2024) 'Fake news detection on social media using ensemble methods', *Computers, Materials & Continua*, Vol. 81, No. 3, pp.4525–4549, DOI: 10.32604/cmc.2024.056291.

Jiang, M., Jing, C., Chen, L., Wang, Y. and Liu, S. (2024) 'An application study on multimodal fake news detection based on Albert-ResNet50 model', *Multimedia Tools and Applications*, January, Vol. 83, No. 3, pp.8689–8706, DOI: 10.1007/s11042-023-15741-y.

Li, L., Jin, T., Lin, W., Jiang, H., Pan, W., Wang, J., Xiao, S., Xia, Y., Jiang, W. and Zhao, Z. (2023) 'Multi-granularity relational attention network for audio-visual question answering', *IEEE Trans. Circuits Syst. Video Technol.*, August, Vol. 34, No. 8, pp.7080–7094, DOI: 10.1109/TCSVT.2023.3264524.

Liu, Q., Wu, J., Wu, S. and Wang, L. (2024) 'Out-of-distribution evidence-aware fake news detection via dual adversarial debiasing', *IEEE Transactions on Knowledge and Data Engineering*, November, Vol. 36, No. 11, pp.6801–6813, DOI: 10.1109/TKDE.2024.3390431.

Ma, Y., Zou, X., Pan, Q., Yan, M. and Li, G. (2024) 'Target-embedding autoencoder with knowledge distillation for multi-label classification', *IEEE Trans. Emerg. Top. Comput. Intell.*, June, Vol. 8, No. 3, pp.2506–2517, DOI: 10.1109/TETCI.2024.3372693.

Narayanan, M.B., Ramesh, A.K., Gayathri, K.S. and Shahina, A. (2023) 'Fake news detection using a deep learning transformer based encoder-decoder architecture', *Journal of Intelligent & Fuzzy Systems*, August, Vol. 45, No. 5, pp.8001–8013, DOI: 10.3233/JIFS-223980.

Preethi, P. and Mamatha, H.R. (2023) 'Region-based convolutional neural network for segmenting text in epigraphical images', *Artif. Intell. Appl.*, September, Vol. 1, No. 2, pp.119–127, DOI: 10.47852/bonviewAIA2202293.

Ren, J., Shi, M., Chen, J., Wang, R. and Wang, X. (2023) 'Hyperspectral image classification using multi-level features fusion capsule network with a dense structure', *Appl. Intell.*, June, Vol. 53, No. 11, pp.14162–14181, DOI: 10.1007/s10489-022-04232-6.

Suryawanshi, S., Goswami, A. and Patil, P. (2024) 'FakeIDCA: fake news detection with incremental deep learning based concept drift adaption', *Multimedia Tools and Applications*, March, Vol. 83, No. 10, pp.28579–28594, DOI: 10.1007/s11042-023-16588-z.

Tang, X., Renteria-Pinon, M. and Tang, W. (2023) 'Dynamic predictive sampling analog to digital converter for sparse signal sensing', *IEEE Trans. Circuits Syst. II Express Briefs*, July, Vol. 70, No. 7, pp.2360–2364, DOI: 10.1109/TCSII.2023.3238279.

Venkateswarlu, B., Shenoi, V.V. and Tumuluru, P. (2023) 'Aquila optimized feedback artificial tree for detection of fake news and impact identification', *Web Intelligence*, November, Vol. 21, No. 2, pp.181–202, DOI: 10.3233/WEB-220046.

Wang, B. and Zhang, S. (2024) 'A multi-reading habits fusion adversarial network for multi-modal fake news detection', *Int. J. Adv. Comput. Sci. Appl.*, July, Vol. 15, No. 7, pp.403–413, DOI: 10.14569/IJACSA.2024.0150740.

Wang, Q., Li, C., Lin, C., Fan, W., Feng, S. and Wang, Y. (2024) 'A news media bias and factuality profiling framework assisted by modeling correlation', *Computers, Materials & Continua*, Vol. 81, No. 2, pp.3351–3369, DOI: 10.32604/cmc.2024.057191.

Wu, Y., Ngai, E.W.T. and Wu, P. (2025) 'Understanding users' news-sharing behaviors: roles of risk perception, believability, fake news awareness and social tie variety', *Ind. Manag. Data Syst.*, January, Vol. 125, No. 2, pp.433–457, DOI: 10.1108/IMDS-01-2024-0039.

Yin, L., Wang, L., Lu, S., Wang, R., Ren, H., AlSanad, A., AlQahtani, S.A., Yin, Z., Li, X. and Zheng, W. (2024) 'AFBNet: a lightweight adaptive feature fusion module for super-resolution algorithms', *CMES-Comp. Model. Eng. Sci.*, May, Vol. 140, No. 3, pp.2315–2347, DOI: 10.32604/cmes.2024.050853.

Zeng, Y. and Xiang, K. (2023) 'Persistence augmented graph convolution network for information popularity prediction', *IEEE Trans. Netw. Sci. Eng.*, November, Vol. 10, No. 6, pp.3331–3342, DOI: 10.1109/TNSE.2023.3258931.

Zhai, Z.L., Zhang, X., Fang, F.F. and Yao, L.Y. (2023) 'Text classification of Chinese news based on multi-scale CNN and LSTM hybrid model', *Multimedia Tools and Applications*, June, Vol. 82, No. 14, pp.20975–20988, DOI: 10.1007/s11042-023-14450-w.

Zhang, D., Shan, G., Lee, M., Zhou, L. and Fu, Z. (2025) 'MT-GPD: a multimodal deep transfer learning model enhanced by auxiliary mechanisms for cross-domain online fake news detection', *Production and Operations Management*, August, Vol. 34, No. 8, pp.2448–2470, DOI: 10.1177/10591478251319686.

Zhang, W., Shen, X., Zhang, H., Yin, Z., Sun, J., Zhang, X. and Zou, L. (2024a) 'Feature importance measure of a multilayer perceptron based on the presingle-connection layer', *Knowl. Inf. Syst.*, January, Vol. 66, No. 1, pp.511–533, DOI: 10.1007/s10115-023-01959-7.

Zhang, X., Peng, Y., Huang, H., Wang, Y., Wang, Q., Lin, Y., Qin, Z. and Gui, G. (2024b) 'SAMS-GNN: self-adaptive multi-scale graph neural network for multi-band spectrum prediction', *IEEE Trans. Cogn. Commun. Netw.*, June, Vol. 11, No. 3, pp.1442–1451, DOI: 10.1109/TCCN.2024.3483202.

Zhu, Y., Zhou, Y., Wei, W. and Zhang, L. (2022) 'Real-time cascading failure risk evaluation with high penetration of renewable energy based on a graph convolutional network', *IEEE Trans. Power Syst.*, September, Vol. 38, No. 5, pp.4122–4133, DOI: 10.1109/TPWRS.2022.3213800.