Journal of Computer Science and Frontier Technologies
ISSN Print: 3104-4204
ISSN Online: 3104-4212
CD ISSN: 3104-4220
https://doi.org/10.63313/JCSFT.9105

A Review of Content Semantics and Modality Relations in Multimodal Fake News Detection

Zenan Qi
School of Information Science and Electrical Engineering (School of Artificial Intelligence), Shandong Jiaotong University, Jinan 250357, China
Email: 2419612481@qq.com

How to cite this paper: Qi, Z. (2026). A review of content semantics and modality relations in multimodal fake news detection. Journal of Computer Science and Frontier Technologies, 4(3), 26–37. ISSN Print: 3104-4204, ISSN Online: 3104-4212. https://doi.org/10.63313/JCSFT.9105
Published: 2026-10-20

Copyright © 2026 by author(s) and Erytis Publishing Limited.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/



Abstract

The key challenge in multimodal fake news detection lies not simply in combining textual and visual information, but in determining whether cues from different modalities jointly support the news narrative. Based on existing literature, this paper categorizes content-driven detection methods into four research directions: content feature representation, semantic associations among samples and events, cross-modal interaction, and content enhancement with adaptive relational reasoning. These methods are comparatively analyzed along three dimensions: information sources, granularity of relationships, and reasoning dependencies. The paper further reviews the label origins, modality composition, and partitioning differences of representative datasets, and clarifies the meanings and applicability of commonly used evaluation metrics. It also discusses several key challenges, including the distinction between semantic consistency and factual veracity, the uncertainty of generative auxiliary features, sample construction risks in contrastive learning, and the efficiency of dynamic fusion models. Rather than simply ranking detection accuracy across different experimental settings, this paper aims to provide a problem-oriented analytical framework to support method selection and experimental comparison.

Keywords

Multimodal Fake News Detection; Content Semantics; Contrastive Learning; Cross-Modal Interaction; Dynamic Fusion

1. Introduction

As social media becomes a major news source, multimodal content combining text and images has become increasingly prevalent. However, misleading cross-modal relationships may distort information, and semantic consistency does not necessarily imply factual authenticity. Therefore, multimodal fake news detection requires both discriminative content representation and effective semantic relationship modeling.

Advances in deep learning and vision-language pretraining have shifted research from simple feature fusion toward four directions: content feature representation,

semantic relationship and contrastive learning, cross-modal interaction, and content augmentation and adaptive reasoning. These approaches exploit complementary features, contextual relationships, semantic alignment, and adaptive fusion to improve detection.

Despite these advances, challenges remain in identifying fine-grained semantic conflicts, constructing reliable semantic relationships, controlling errors in generated information, and balancing detection performance with computational efficiency. Dataset and evaluation differences further complicate method comparisons.

Existing surveys primarily focus on detection architectures and fusion strategies[1-4], offering limited comparisons of information sources, semantic modeling objectives, and reasoning dependencies. To address this, this review systematically examines representative methods across the four directions, comparing their mechanisms, performance, and limitations. It also discusses datasets, evaluation settings, and future research challenges.

The remainder of this paper is organized as follows. Section 2 introduces the analytical framework. Section 3 reviews representative methods and their comparisons. Section 4 presents datasets and evaluation metrics. Section 5 discusses challenges and future directions, and Section 6 concludes the paper.

2. Related Work

2.1. Scope of Multimodal Fake News Detection

Multimodal fake news detection aims to identify false or misleading news using textual, visual, and auxiliary information. Based on information sources, existing approaches can be categorized into content-driven, propagation-driven, and evidence-driven methods. Content-driven methods rely on news text, images, and derived features, as exemplified by BCMF[5]. Propagation-driven methods exploit user interactions and propagation structures, such as DPSG[6], while evidence-driven methods incorporate external evidence for authenticity assessment, as demonstrated by MUSER[7]. These approaches may overlap but differ in their data requirements and applicability.

This review focuses on content-driven detection using news text and associated images as primary inputs. Linguistic styles, local visual features, and generated image descriptions are treated as content-derived cues rather than verified evidence. Additional information, including related news contexts, user interactions, propagation networks, and external evidence, should be explicitly identified when comparing methods.

Importantly, cross-modal semantic consistency does not necessarily imply factual authenticity, nor does inconsistency establish falsity. Therefore, cross-modal relationships should serve as supplementary discriminative cues rather than substitutes for fact verification. The key challenge is to extract informative features and identify semantic relationships relevant to authenticity assessment.

2.2. Analytical Framework for Multimodal Fake News Detection

From a modeling perspective, content-driven multimodal fake news detection can be analyzed at three interrelated levels: feature representation, semantic relationship modeling, and fusion-based decision-making. These levels provide a framework for comparing different approaches rather than defining a fixed processing pipeline.

The feature representation level extracts discriminative textual and visual information through semantic encoding, multi-view representations, and multi-scale visual modeling. For example, MViR captures complementary visual-semantic features[8]. The main challenge is to preserve informative features while reducing redundancy and noise.

The semantic relationship level models associations and discrepancies among news samples, event contexts, and cross-modal information. Methods such as contrastive learning and semantic alignment capture relationships at different granularities. CSFND, for instance, employs contextual semantic representation learning[9]. The key challenge is to construct reliable semantic relationships without introducing misleading associations.

The fusion-based decision-making level integrates multimodal features through attention mechanisms, relationship-aware fusion, and adaptive expert models. MIMoE-FND uses modality interaction and a mixture-of-experts mechanism for adaptive fusion[10]. However, reliable fusion and computational efficiency remain important considerations.

In addition, information sources serve as a complementary analytical dimension, distinguishing original news content, derived features, related corpora, and external information. The overall framework is summarized in Table 1.

Table 1. Analytical dimensions of multimodal fake news detection methods.

Analytical dimension	Core question	Main technical approaches
Feature representation	How can discriminative textual and visual information be extracted?	Multi-view textual representations, multi-scale visual encoding, content-derived features
Semantic relationships	How can semantic associations and discrepancies be modeled?	Event contexts, contrastive learning, multi-granularity semantic alignment
Fusion-based decision-making	How can multimodal information be integrated for detection?	Cross-modal attention, relationship-aware fusion, dynamic expert mechanisms
Information sources	What inputs and auxiliary information are required?	Original text and images, derived information, related corpora, propagation information, and external evidence

The first three dimensions characterize modeling objectives and technical approaches, while information sources clarify data dependencies. Accordingly, Section 3 reviews four complementary research directions: content feature representation, semantic relationship and contrastive learning, cross-modal interaction, and content augmentation and adaptive reasoning.

3. Major Methodological Approaches to Content-Driven Detection

3.1. Content Feature Representation

Content feature representation aims to extract discriminative information from news text and images. Early methods focused on lexical, stylistic, and statistical features, as exemplified by ConFake[11]. Deep learning methods further enhance contextual representations, with CMGN employing text graph neural networks and feature fusion to model textual relationships [12].

Beyond semantic features, SSM integrates linguistic styles, semantic similarity, and visual cues for detection[13]. However, SheepDog highlights the vulnerability of style-dependent methods to LLM-based attacks[14], suggesting that stylistic features should serve as supplementary cues.

Visual representation has evolved toward multi-view and multi-scale modeling. MViR captures complementary visual-semantic features[8], while MSACA combines multi-scale visual features with cross-modal semantic alignment [15]. Although these approaches enrich content representations, they may introduce feature redundancy and computational overhead.

Overall, effective content representation requires balancing discriminative information, feature redundancy, and model complexity. Table 2 compares representative methods in terms of their mechanisms, performance, and limitations.

Table 2. Comparison of representative content feature representation method.

Method	Information Used	Key Mechanism	Datasets	Main Findings	Potential Limitations
ConFake[11]	Text	80 handcrafted features, word embeddings, and ML classifiers	Kaggle, McIntire, Reuters, BuzzFeed, PolitiFact	97.31% Acc (reported)	Handcrafted feature dependence; no visual information
CMGN[12]	Text, auxiliary text	Text GNN, RWKV MLP-Mixer, and cross-feature fusion	LIAR, FA-KES, IFND, CHEF	Outperformed compared baselines	Depends on auxiliary text and graph quality; no visual information
SSM[13]	Text, images	Stylometric features, semantic consistency, and forgery detection	FNDD, TI-CNN, FNSDS	Acc: 96.74%, 96.50%, 96.90%	Complex architecture; depends on reliable forgery and consistency cues
MViR[8]	Text, images	Pyramid dilated convolution and multi-view semantic fusion	Weibo, GossipCop	Acc: 92.4%, 89.5%	Feature redundancy; 61.9% fake-class recall on GossipCop
MSACA[15]	Text, images	Multi-scale semantic alignment and cross-modal attention	Weibo, GossipCop	Acc: 90.3%, 88.7%	Depends on alignment quality; potentially sensitive to misleading images

3.2. Semantic Relationships and Contrastive Learning

Semantic relationship modeling extends beyond individual news items to capture contextual and cross-sample information. NEP constructs external news environments to extract contextual cues[16], while CSFND learns contextual semantic representations to distinguish semantically similar news with different authenticity labels[9]. Such approaches may require additional news resources beyond the original content.

Event-Radar further models event-level semantic relationships and cross-modal inconsistencies through event-driven multi-view learning[17]. Contrastive learning enhances discriminative representations by exploiting sample relationships. CCGN combines consistency-based contrastive learning with graph modeling[18], while Zeng et al. address biases in multimodal inconsistency learning[19]. DCLR further mitigates modality imbalance through dynamic contrastive learning[20].

As summarized in Table 3, these methods improve semantic relationship modeling through contextual learning, event-level reasoning, and contrastive strategies. However, their effectiveness depends on reliable relationship construction and sample selection. Since semantic similarity and event relatedness do not necessarily imply identical authenticity labels, distinguishing meaningful authenticity cues from superficial semantic associations remains essential for robust detection.

Table 3. Comparison of semantic relationship modeling and contrastive learning methods.

Method	Information Used	Key Mechanism	Datasets	Main Findings	Potential Limitations
NEP [16]	Text, external news	Macro-/micro-level news environments; popularity and novelty modeling	Chinese (39,066), English (6,483)	Acc: 83.1%, 72.8%	External news dependence; weaker English performance
CSFND[9]	Text, images	Local context learning, global semantic features, and contextual classification	Weibo	Average Acc improvement: 2.5%	Context quality dependence; uncertain robustness to unseen semantics
Event-Radar [17]	Text, images	Event graph construction and credibility-aware multi-view fusion	Twitter, Weibo, PHEME	Acc: 92.8%, 91.9%, 90.1%	Depends on event graph accuracy and view credibility
CCGN [18]	Text, images	Visual GNN, consistency contrastive learning, and cross-modal attention	Weibo, Twitter	Acc: 90.8%, 90.6%	Alignment dependence; potential difficulty with subtle factual conflicts
MMICF[19]	Text, images	Causal reasoning, modality unification, and inconsistency contrastive learning	Fakeddit, Weibo	Acc: 96.91%, 96.57%	Depends on causal modeling; difficulty with implicit inconsistencies
DCLR [20]	Text, images	Clustering-purity monitoring, re-initialization, and dynamic contrastive learning	Weibo, Twitter	Not specified	Training overhead; imbalance estimation and threshold sensitivity

3.3. Cross-Modal Interaction

Cross-modal interaction aims to capture complementary information and semantic discrepancies between text and images. Research has evolved from simple feature fusion toward attention-based interaction. BCMF employs bidirectional cross-modal fusion[5], while HMCAN uses hierarchical contextual attention to model textual semantics and cross-modal relationships[21]. However, such interactions may also introduce irrelevant information.

Vision–language pretraining further enhances semantic alignment. MSACA combines multi-scale alignment with cross-modal attention[15], while SARD integrates CLIP-based contrastive learning and semantic alignment[22]. Notably, SARD also relies on social contextual information, distinguishing it from purely content-driven methods.

Beyond alignment, MIAN employs inverse attention to capture cross-modal discrepancies[23], while Tao et al. introduce consistency suppression to reduce potentially misleading alignment signals[24]. These approaches emphasize the importance of modeling semantic conflicts rather than similarity alone.

Moreover, MMDFND incorporates domain-aware modeling to address variations in cross-modal dependencies[25]. Although these methods improve interaction flexibility, they may increase model complexity and computational costs. Table 4 summarizes representative cross-modal interaction methods and their performance and limitations.

Table 4. Comparison of representative cross-modal interaction methods.

Method	Information Used	Key Mechanism	Datasets	Main Findings	Potential Limitations
BCMF[5]	Text, images	Bidirectional cross-modal fusion	Weibo, Twitter, PolitiFact, GossipCop	Acc gains: 2.2–4.9 pp	Potential cross-modal noise propagation
HMCAN[21]	Text, images	BERT/ResNet and hierarchical contextual attention	Weibo, Twitter, PHEME	Acc: 88.5%, 89.7%, 88.1%	Limited explicit factual conflict modeling
MSACA[15]	Text, images	Multi-scale semantic alignment and cross-modal attention	Weibo, GossipCop	Acc: 90.3%, 88.7%	Alignment dependence; sensitivity to misleading images
SARD [22]	Text, images, social context	CLIP contrastive learning, semantic alignment, N-HIN fusion	PHEME, GossipCop, additional dataset	Avg. gains: 2.89% Macro-F1; 2.13% Acc	Social context dependence; complex alignment

Method	Information Used	Key Mechanism	Datasets	Main Findings	Potential Limitations
MIAN [23]	Text, images	Hierarchical learning, co-attention, inverse attention	Weibo17, Weibo21, GossipCop, PolitiFact	Acc: 93.6%, 93.8%, 92.3%, 95.3%	Depends on reliable inconsistency extraction
MCSF [24]	Text, images	CLIP alignment, consistency suppression, VAE reconstruction	Weibo, Twitter, Weibo21	Acc: 92.6%, 83.9%, 92.3%	Consistency estimation dependence; no explicit fact verification
MMDFND[25]	Text, images, domain information	Domain-aware features, Stepwise Pivot Transformers, adaptive normalization	Weibo, Weibo21	Acc: 93.4%, 93.9%	Domain dependence; uncertain unseen-domain generalization

3.4. Content Augmentation and Adaptive Reasoning

Content augmentation enriches the semantic representations of news through visual descriptions, natural language transformation, and generative models. UNITE-FND converts visual content into structured textual descriptions for text-based classification[26], while NLIN transforms images and related knowledge into natural language representations for contextual reasoning[27]. SEER combines visual semantic enhancement with emotional reasoning[28]. Entity-enhanced methods further integrate multimodal clues associated with news entities[29]. Recent studies also exploit LLM-generated content and negative reasoning to support detection, explanation, and discriminative learning [30, 31]. However, generated information may contain inaccuracies or reasoning errors, and its reliability cannot be guaranteed by linguistic coherence alone. Methods incorporating external knowledge, such as NLIN, also require explicit consideration of their additional information dependencies.

Adaptive reasoning methods dynamically adjust multimodal information integration according to sample-specific relationships. MIMoE-FND employs modality interaction assessment and hierarchical mixture-of-experts mechanisms[10], while GAMED combines multiple expert networks with feature decoupling for adaptive fusion[32]. Although these methods offer flexible decision-making, their performance depends on effective expert specialization and routing, potentially increasing model complexity and inference costs.

Overall, content augmentation and adaptive reasoning enrich semantic information and enable flexible multimodal decision-making, but differ in information sources, reasoning mechanisms, and computational requirements. Table 5 compares representative methods in terms of their core mechanisms, experimental results, and potential limitations.

Table 5. Comparison of content augmentation and adaptive reasoning methods.

Method	Information Used	Key Mechanism	Datasets	Main Findings	Potential Limitations
UNITE-FND [26]	Generated visual descriptions	Multimodal-to-text translation and text classification	Uni-Fakedit-55k	92.52% Acc in binary classification	Information loss and dependence on description generation
NLIN[27]	Image descriptions, external knowledge	Natural language unification and contextual inference	CFND and other evaluation settings	Reported improved reasoning performance	External knowledge dependence and conversion errors
SEER[28]	Image captions, emotional cues	Semantic enhancement and expert-based emotional reasoning	Weibo, Twitter	Reported improvements over baselines	Caption uncertainty and emotion-related bias
LLM-MFEFND [30]	LLM-generated background information and comments	Hierarchical progressive multimodal fusion	Weibo21, FineFake	94.5% and 81.1% Acc, respectively	Generation reliability and additional inference costs
MIMoE-FND[10]	Text, images, modality relations	Interaction gating and hierarchical mixture of experts	Weibo, Weibo21, GossipCop	Reported improved adaptive fusion	Routing reliability and expert utilization
GAMED [32]	Text, images, semantic knowledge	Feature decoupling and knowledge-adaptive multi-expert fusion	Fakeddit, Yang	Reported improvements over comparison methods	Expert specialization and computational overhead

3.5. Applicability Boundaries of Related Approaches

Beyond content-driven detection, evidence retrieval, propagation modeling, and cross-domain transfer provide complementary information. MUSER uses multi-step evidence retrieval[7], DPSG models social propagation patterns[6], and Yang et al. explore domain transfer with user engagement information[33]. However, these approaches often depend on external data and specific settings, requiring careful consideration of information sources and evaluation conditions.

Overall, the four research directions provide complementary perspectives on multimodal fake news detection. Their effectiveness is influenced by information sources, relationship granularity, and model complexity, which also affect subsequent dataset and evaluation analyses.

4. Commonly Used Datasets and Evaluation Metrics

4.1. Commonly Used Datasets and Their Characteristics

Existing multimodal fake news datasets differ in modality composition, label sources, and evaluation settings. Among social media datasets, Weibo provides Chinese text, images, and social information[34,35], while Twitter/MediaEval focuses on misleading multimedia use[36,37]. Fakeddit contains over one million Reddit samples with distant supervision labels for binary, three-class, and six-class classification[38].

For news-based detection, FakeNewsNet combines news content and propagation information through PolitiFact and GossipCop[39]. Yang/TI-CNN provides approximately 20,000 multimodal news samples[40], while NewsBag and NewsBag++ contain genuine and satirical news for studying class imbalance[41]. BuzzFeed-Webis includes 1,627 fact-checked articles [43].

Task-specific datasets include ReCOVery, containing 2,029 COVID-19 news articles labeled mainly by publisher credibility[42], and PHEME, supporting rumor detection, veracity assessment, and stance analysis[44].

Overall, differences in modality availability, label definitions, and auxiliary information affect experimental comparability, highlighting the need to specify dataset versions, annotation criteria, and input configurations.

4.2. Evaluation Metrics and Experimental Comparability

Multimodal fake news detection is commonly evaluated using Accuracy, Precision, Recall, and F1-score. For imbalanced datasets, class-wise metrics and Macro-F1 provide a more comprehensive assessment, while ROC-AUC measures discrimination across classification thresholds. Consistent positive-class definitions and metric calculations are essential for reliable comparisons.

Experimental settings also affect reported performance. Differences in dataset versions, preprocessing, and data splits may alter data distributions, while overlapping events or repeated images across training and test sets can lead to overly optimistic results. Event-disjoint evaluation, temporal splits, and cross-dataset testing offer additional insights into generalization.

Furthermore, methods may differ in their use of auxiliary information, such as generated captions, related news, user interactions, and external evidence. These dependencies, together with model complexity and inference costs, should be reported to ensure fair comparisons. Overall, reliable evaluation requires consistent metrics, transparent experimental protocols, and explicit documentation of input modalities and auxiliary information, providing a basis for comparing information sources, semantic relationships, and fusion strategies.

5. Challenges and Future Research Directions

Despite recent advances, multimodal fake news detection faces several challenges. Global semantic alignment may overlook fine-grained conflicts involving entities and events, requiring more precise cross-modal modeling. Generated descriptions and LLM-based augmentation may introduce errors, highlighting the need for reliability assessment and uncertainty-aware reasoning. Dataset biases and distribution shifts limit generalization, calling for cross-event, cross-domain, and temporal evaluations. Meanwhile, complex fusion mechanisms increase computational costs, motivating lightweight architectures and efficient inference. Future research should address these challenges to improve detection reliability, generalization, and practical

applicability.

6. Conclusion

This review examines multimodal fake news detection through four directions: content feature representation, semantic relationship and contrastive learning, cross-modal interaction, and content augmentation and adaptive reasoning. Representative methods are compared in terms of information sources, semantic granularity, and reasoning dependencies, alongside dataset and evaluation differences. Although multimodal information provides complementary cues, complex fusion does not necessarily ensure reliable authenticity assessment. Future research should emphasize fine-grained semantic conflict detection, auxiliary information reliability, cross-domain generalization, and computational efficiency.

References

[1] LEE J, CHA J, PARK M, et al. A survey of fake news detection: A comprehensive review of multimodal approaches[J]. Data & Knowledge Engineering, 2026, 165: 102622.

[2] NASSER M, ARSHAD N I, ALI A, et al. A systematic review of multimodal fake news detection on social media using deep learning models[J]. Results in Engineering, 2025, 26: 104752.

[3] ZIDAN M, SLEEM A, NABIL A, et al. Multimodal fake news detection: A survey of text and visual content integration methods[J]. International Journal of Computers and Informatics (Zagazig University), 2025, 7: 13 – 25.

[4] ÇETINKAYA A, AGHAYEV H. Evolution of Fake-News Detection: Conventional, Automatic, and AI-Based Methods[M]//Reshaping Journalism and Communications With AI. IGI Global Scientific Publishing, 2026: 165 – 188.

[5] YU C, MA Y, AN L, et al. BCMF: A bidirectional cross-modal fusion model for fake news detection[J]. Information Processing & Management, 2022, 59(5): 103063.

[6] JING C, GAO H, ZHANG X, et al. DPSG: Dynamic propagation social graphs for multi-modal fake news detection[J]. Information Fusion, 2025, 113: 102595.

[7] LIAO H, PENG J, HUANG Z, et al. Muser: A multi-step evidence retrieval enhancement framework for fake news detection[C]//Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2023: 4461 – 4472.

[8] LIANG H, SU X, WANG J, et al. MViR: Multi-View Visual-Semantic Representation for Fake News Detection[C]//2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2026: 12192 – 12196.

[9] PENG L, JIAN S, KAN Z, et al. Not all fake news is semantically similar: Contextual semantic representation learning for multimodal fake news detection[J]. Information Processing & Management, 2024, 61(1): 103564.

[10] LIU Y, LIU Y, LI Z, et al. Modality interactive mixture-of-experts for fake news detection[C]//Proceedings of the ACM Web Conference 2025. 2025: 5139 – 5150.

[11] JAIN M K, GOPALANI D, MEENA Y K. ConFake: Fake news identification using content based features[J]. Multimedia Tools and Applications, 2024, 83(3): 8729 – 8755.

[12] CUI S, DUAN K, MA W, et al. CMGN: Text GNN and RWKV MLP-mixer combined with cross-feature fusion for fake news detection[J]. Neurocomputing, 2025, 633: 129811.

[13] NADEEM M I, AHMED K, ZHENG Z, et al. SSM: Stylometric and semantic similarity oriented multimodal fake news detection[J]. Journal of King Saud University - Computer and Information Sciences, 2023, 35(5): 101559.

[14] WU J, GUO J, HOOI B. Fake news in sheep's clothing: Robust fake news detection against

LLM-empowered style attacks[C]//Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2024: 3367–3378.

[15] WANG J, ZHANG H, LIU C, et al. Fake news detection via multi-scale semantic alignment and cross-modal attention[C]//Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2024: 2406–2410.

[16] SHENG Q, CAO J, ZHANG X, et al. Zoom out and observe: News environment perception for fake news detection[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2022: 4543–4556.

[17] MA Z, LUO M, GUO H, et al. Event-Radar: Event-driven multi-view learning for multimodal fake news detection[C]//Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2024: 5809–5821.

[18] CUI S, DUAN K, MA W, et al. CCGN: Consistency contrastive-learning graph network for multi-modal fake news detection[J]. Multimedia Systems, 2025, 31(2): 119.

[19] ZENG Z, WU M, LI G, et al. Correcting the bias: Mitigating multimodal inconsistency contrastive learning for multimodal fake news detection[C]//2023 IEEE International Conference on Multimedia and Expo (ICME). 2023: 2861–2866.

[20] WANG Z, JIANG C, WANG K. DCLR: Mitigating modality imbalance in multimodal fake news detection via dynamic contrastive learning with re-initialization[J]. Multimedia Systems, 2026, 32(1): 45.

[21] QIAN S, WANG J, HU J, et al. Hierarchical multi-modal contextual attention network for fake news detection[C]//Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2021: 153–162.

[22] YAN F, ZHANG M, WEI B, et al. SARD: Fake news detection based on CLIP contrastive learning and multimodal semantic alignment[J]. Journal of King Saud University - Computer and Information Sciences, 2024, 36(8): 102160.

[23] ZHANG T, YU E, SHAO Y, et al. Multimodal inverse attention network with intrinsic discriminant feature exploitation for fake news detection[EB/OL]. arXiv:2502.01699, 2025. https://arxiv.org/abs/2502.01699.

[24] TAO Z, ZHAO R, SHI X, et al. Multimodal consistency suppression factor for fake news detection[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2025, 21(11): Article 319, 1–19.

[25] TONG Y, LU W, ZHAO Z, et al. MMDFND: Multi-modal multi-domain fake news detection[C]//Proceedings of the 32nd ACM International Conference on Multimedia. 2024: 1178–1186.

[26] MUKHERJEE A, GHOSH S. UNITE-FND: Reframing multimodal fake news detection through unimodal scene translation[EB/OL]. arXiv:2502.11132, 2025. https://arxiv.org/abs/2502.11132.

[27] ZHANG Q, LIU J, ZHANG F, et al. Natural language-centered inference network for multi-modal fake news detection[C]//Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI). 2024: 2542–2550.

[28] ZHU P, JING Y, CHENG L, et al. SEER: Semantic enhancement and emotional reasoning network for multimodal fake news detection[C]//2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC). 2025: 1768–1774.

[29] QI P, CAO J, LI X, et al. Improving fake news detection by using an entity-enhanced framework to fuse diverse multimodal clues[C]//Proceedings of the 29th ACM International Conference on Multimedia. 2021: 1212–1220.

[30] WANG L, LI X, ZHOU B, et al. Multimodal fusion with LLM content via hierarchical progressive transformer for explainable fake news detection[J]. Information Processing & Management, 2026, 63(5): 104700.

[31] ZHANG C, FENG Z, ZHANG Z, et al. Is LLMs hallucination usable? LLM-based negative reasoning for fake news detection[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2025, 39(1): 1031–1039.

[32] SHEN L, LONG Y, CAI X, et al. GAMED: Knowledge adaptive multi-experts decoupling for multimodal fake news detection[C]//Proceedings of the Eighteenth ACM International Conference on Web Search and Data Mining. 2025: 586–595.

[33] YANG X, WANG Y, ZHANG X, et al. A macro- and micro-hierarchical transfer learning framework for cross-domain fake news detection[C]//Proceedings of the ACM Web Conference 2025. 2025: 5297–5307.

[34] JIN Z, CAO J, GUO H, et al. Multimodal fusion with recurrent neural networks for rumor detection on microblogs[C]//Proceedings of the 25th ACM International Conference on Multimedia. 2017: 795–816. https://doi.org/10.1145/3123266.3123454.

[35] CAO J, SHENG Q, QI P, et al. False news detection on social media[EB/OL]. arXiv:1908.10818, 2019. https://arxiv.org/abs/1908.10818.

[36] BOIDIDOU C, ANDREADOU K, PAPADOPOULOS S, et al. Verifying multimedia use at MediaEval 2015[C]//Working Notes Proceedings of the MediaEval 2015 Workshop. CEUR Workshop Proceedings, 2015, 1436.

[37] BOIDIDOU C, PAPADOPOULOS S, DANG-NGUYEN D T, et al. Verifying multimedia use at MediaEval 2016[C]//Working Notes Proceedings of the MediaEval 2016 Workshop. CEUR Workshop Proceedings, 2016, 1739.

[38] [38] NAKAMURA K, LEVY S, WANG W Y. Fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection[C]//Proceedings of the Twelfth Language Resources and Evaluation Conference. 2020: 6149–6157.

[39] SHU K, MAHUDESWARAN D, WANG S, et al. FakeNewsNet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media[J]. Big Data, 2020, 8(3): 171–188. https://doi.org/10.1089/big.2020.0062.

[40] YANG Y, ZHENG L, ZHANG J, et al. TI-CNN: Convolutional neural networks for fake news detection[EB/OL]. arXiv:1806.00749, 2018. https://arxiv.org/abs/1806.00749.

[41] JINDAL S, SOOD R, SINGH R, et al. NewsBag: A multimodal benchmark dataset for fake news detection[C]//CEUR Workshop Proceedings. 2020, 2560: 138–145.

[42] ZHOU X, MULAY A, FERRARA E, et al. ReCOVery: A multimodal repository for COVID-19 news credibility research[C]//Proceedings of the 29th ACM International Conference on Information & Knowledge Management. 2020: 3205–3212. https://doi.org/10.1145/3340531.3412880.

[43] POTTHAST M, KIESEL J, REINARTZ K, et al. A stylometric inquiry into hyperpartisan and fake news[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2018: 231–240. https://doi.org/10.18653/v1/P18-1022.

[44] ZUBIAGA A, LIAKATA M, PROCTER R, et al. Analysing how people orient to and spread rumours in social media by looking at conversational threads[J]. PLOS ONE, 2016, 11(3): e0150989. https://doi.org/10.1371/journal.pone.0150989.