International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(1s) 2026 pp. 1-17
https://doi.org/10.70917/ijcisim-2026-2006
Article

Trust-Aware Cross-Modal Inconsistency Fusion for Text-Image Fake News Detection in Social Media

Naeem Akhtar1, Anurag Rana2

[1,2]Yogananda School of AI, Computers and Data Sciences, Shoolini University Himachal Pradesh, India
Corresponding Author mail: akhtar1992naeem@gmail.com

Abstract: Multimodal fake-news detection is important because misleading social media posts often combine textual claims with visual content to construct deceptive narratives. This study proposes T-CAMIF, a Trust-Aware Cross-Modal Inconsistency Fusion framework for detecting fake news in paired text-image social media posts. The framework models three complementary evidence sources: textual semantics, visual representations, and cross-modal inconsistency. Textual features were extracted using a Sentence Transformer/MiniLM-based encoder, while visual features were extracted using a MobileNetV2-based encoder. Cross-modal inconsistency was represented using cosine similarity, cosine distance, Euclidean distance, absolute feature difference, and element-wise product. Evidence-level predictions were integrated through a trust-aware fusion gate that incorporated expert probabilities, confidence scores, uncertainty indicators, and disagreement measures. The framework was evaluated on a public Twitter-Weibo multimodal fake-news dataset using accuracy, precision, recall, F1-score, macro-F1, and AUC. Experimental results showed that T-CAMIF achieved the strongest in-domain performance on the combined dataset, with an accuracy of 0.9061, F1-score of 0.9095, macro-F1 of 0.9059, and AUC of 0.9709. It outperformed unimodal, inconsistency-only, simple concatenation, and static full-fusion baselines. Ablation and robustness analyses confirmed the contribution of cross-modal inconsistency and trust-aware fusion, while also showing reduced performance under missing modalities and mismatched image-text pairs. Cross-platform evaluation further indicated that platform variation remains a key challenge. These findings suggest that trust-aware modelling of text-image relationships can improve multimodal fake-news detection and motivate future work on domain adaptation, stronger vision-language encoders, and uncertainty-aware multimodal learning.

Keywords: Multimodal fake-news detection; Trust-aware fusion; Cross-modal inconsistency; Text-image classification; Uncertainty-aware learning.

1. Introduction

Multimodal fake news detection has become an important research problem because deceptive social media content combines textual claims with visual evidence. Posts often use images, captions, screenshots, and memes to create narratives whose falsity may not be identifiable from either modality alone. Deception may emerge from cross-modal inconsistency: an authentic image may be paired with a false caption, an unrelated photograph may support a fabricated claim, or a credible visual may distort an event. Therefore, robust detection requires models that analyse textual meaning, visual evidence, and their semantic alignment. Prior studies show that modelling multimodal consistency and vision-language interaction improves fake-news and disinformation detection because weakly aligned or contradictory text-image pairs can reveal deception cues [1], [2].

The scale and speed of social media further complicate misinformation detection. Unlike traditional news environments, social platforms allow rapid content generation, modification, and circulation with limited editorial control. Early computational methods mainly relied on textual features such as lexical patterns, sentiment, writing style, and semantic representations. However, recent work shows that social-media misinformation is increasingly multimodal and requires models capable of analysing textual, visual, and contextual information [3]. Systematic reviews also report that deep learning has improved detection performance, although challenges remain in modality

alignment, explainability, robustness, and generalization [4]. These issues are important because misleading meaning is often produced through text-image interaction rather than through a single modality [5].

Fake news detection is socially significant because misinformation can affect public health, electoral behaviour, financial decision-making, social stability, and institutional trust. From a data-mining perspective, fake news involves content characteristics, source credibility, user engagement, propagation patterns, and social context [6]. It is also a multidimensional phenomenon involving false knowledge, deceptive intention, social influence, and cognitive vulnerability [7]. Moreover, misinformation spreads through interactions among human behaviour, platform design, and information ecosystems [8]. These characteristics require detection systems that process heterogeneous multimodal content, especially when visual evidence appears credible but is semantically disconnected from the accompanying claim.

A central limitation in existing research is that many models prioritize classification accuracy without sufficiently assessing multimodal evidence trustworthiness. In posts, the text may appear plausible and the image may appear authentic, but their combination may still create a misleading interpretation. A reliable model should determine whether modalities are supportive, contradictory, irrelevant, or uncertain. Earlier studies identified challenges such as ambiguous evidence, evolving misinformation patterns, limited labelled data, and difficulty in modelling deceptive intent [9]. Although deep learning enables automatic feature learning, many models remain constrained by black-box decision-making and limited interpretability [10]. This creates a need for methods that explicitly model text-image inconsistency while estimating evidence reliability.

This study addresses this gap by proposing T-CAMIF: Trust-Aware Cross-Modal Inconsistency Fusion, a framework for fake-news detection in paired text-image posts. The argument is that multimodal detection should not simply merge text and image features using fixed or equal weights. Instead, it should evaluate semantic alignment, evidential reliability, uncertainty, and disagreement before final prediction. This trust-aware perspective is important because social-media content often contains noisy captions, irrelevant images, manipulated visuals, or emotionally charged claims. Explainable artificial intelligence emphasizes transparent outputs in socially sensitive domains [11], while uncertainty-aware deep learning highlights confidence estimation under ambiguity, conflicting evidence, or incomplete information [12].

Benchmark datasets support multimodal fake-news detection. Twitter-based datasets have enabled analysis of how images and text jointly contribute to misleading posts [13]. Fakeddit expanded this direction through a benchmark for fine-grained multimodal misinformation classification [14]. FakeNewsNet integrated news content, social context, and spatiotemporal information for fake-news research [15]. These resources show that single-modality analysis is insufficient and that effective detection requires modelling relationships among textual claims, visual evidence, and contextual reliability.

Despite these advances, four gaps remain. First, many multimodal approaches emphasize feature fusion but do not sufficiently distinguish supportive, contradictory, or irrelevant text-image relationships. Second, several models improve accuracy but provide limited explanation for predictions. Third, uncertainty and trust estimation remain underdeveloped in multimodal fake-news detection. Fourth, platform-specific variation continues to limit generalization. To address these gaps, this study develops a trust-aware multimodal inconsistency fusion framework that models cross-modal mismatch, estimates evidence reliability, and adaptively fuses multiple expert predictions.

The main contributions are fourfold. First, the study proposes a trust-aware cross-modal inconsistency fusion framework for text-image fake-news detection. Second, it designs separate text, image, and inconsistency evidence streams to preserve modality-specific and cross-modal signals. Third, it introduces a trust-aware fusion mechanism using expert probabilities, confidence, uncertainty, and disagreement indicators. Fourth, it validates the framework through baseline comparison, ablation, robustness testing, cross-platform evaluation, multi-seed validation, and paired bootstrap comparison.

The research objectives are:

1. to develop a trust-aware cross-modal inconsistency fusion framework for fake-news detection
2. to compare T-CAMIF with unimodal, inconsistency-only, and conventional multimodal fusion baselines using standard metrics
3. to validate the framework through ablation, robustness testing, cross-platform evaluation, multi-seed validation, and paired bootstrap comparison

2. Literature Review

Recent research on multimodal fake news detection has moved beyond conventional text-only classification toward approaches that model uncertainty, modality interaction, semantic inconsistency, and explainability. Uncertainty-aware disentangled representation learning has been used to separate reliable and uncertain multimodal signals, which is important because social media posts frequently contain ambiguous or conflicting evidence [16]. Cross-modal content correlation has also been shown to provide useful cues for identifying misinformation by examining the relationship between textual claims and visual content [17]. Explainable clue-mining methods further demonstrate that modality-common, modality-specific, and inconsistent evidence can strengthen multimodal fake news detection [18]. Similarly, fuzzy-based multimodal reasoning highlights the value of transparent decision mechanisms in improving interpretability and classification reliability [19]. These studies show that uncertainty and explainability are central to robust multimodal fake news detection; however, they often treat uncertainty estimation, cross-modal clue extraction, and interpretability as separate modelling tasks rather than integrating them into a unified trust-aware fusion mechanism.

Fusion strategy remains a major challenge in multimodal fake news detection. Progressive fusion networks have shown that staged integration of textual and visual representations can be more effective than direct concatenation [20]. Other studies have identified the neutralization effect in modality fusion, where useful information from one modality may be weakened by noisy or dominant features from another modality [21]. Balanced multimodal learning with hierarchical fusion addresses this issue by controlling modality imbalance and combining feature-level and modality-level information [22]. Multi-view mutual learning further supports collaborative learning from complementary multimodal perspectives to improve robustness [23]. These works indicate that simple text-image combination is insufficient for reliable detection. Effective fusion must preserve complementary evidence, reduce modality dominance, and adapt to the reliability of each modality. Nevertheless, many existing fusion strategies still do not explicitly estimate how trustworthy each evidence source is for an individual sample.

A related research direction focuses on cross-modal interaction, ambiguity, and representation granularity. Multi-grained information fusion shows that misinformation can be better represented by combining fine-grained and coarse-grained multimodal cues [24]. Cross-modal ambiguity learning is particularly relevant because text-image relations in fake news are often uncertain rather than clearly matched or mismatched [25]. Earlier foundational models, including EANN, MVAE, SpotFake, and SpotFake+, established the importance of multimodal representation learning, event-bias reduction, joint text-image encoding, and transfer learning for fake news detection [26]–[29]. Collectively, these studies confirm that multimodal fake news detection depends not only on the availability of text and image, but also on the quality of their semantic relationship. Cross-modal ambiguity, event-specific bias, and representation mismatch remain persistent challenges, which suggests that detection models should explicitly evaluate whether modalities are semantically aligned, weakly related, or contradictory.

Recent studies have further expanded the field through event-aware, knowledge-augmented, manipulation-sensitive, and semantic-enhancement approaches. Event-driven multi-view learning demonstrates the importance of event-level information in social media misinformation detection [30]. Knowledge-augmented large vision-language models show that external knowledge can support reasoning when textual and visual evidence alone is insufficient [31]. Research on manipulated images emphasizes that visual authenticity should be considered alongside text-image alignment because deceptive posts may depend on either image manipulation or misleading image use [32]. Bootstrapped multi-view representations and multiscale semantic enhancement also support the shift from single fused representations toward complementary and structured multimodal learning [33], [34]. These studies broaden multimodal fake news detection by incorporating event context, knowledge, visual manipulation, and semantic enrichment, but they also indicate that models may remain sensitive to event-specific or platform-specific patterns.

Adaptive and domain-sensitive learning has therefore become increasingly important. Dynamic hierarchical memory and mixture-of-experts models suggest that expert routing can help capture complex multimodal misinformation patterns [35]. Multi-granularity consistency modelling reinforces the need to analyse text-image consistency at different semantic levels rather than as a single global feature [36]. Domain-aware multimodal multi-view learning further shows that domain shift remains a major challenge when detection models are transferred across platforms, events, or topics [37]. These works support the need for adaptive models that separate stable misinformation cues from domain-dependent signals. However, domain-aware and expert-based approaches still require stronger integration with cross-modal inconsistency and trust estimation so that the model can determine which evidence source should be emphasized under different conditions.

Attention-based and inconsistency-oriented methods provide additional foundations for the present study. Cross-modal attention and residual learning enhance interaction between textual and visual representations [38]. Multimodal fusion with inconsistency reasoning directly connects text-image mismatch with explainable fake news detection [39]. Similarity-aware multimodal prompt learning shows that similarity modelling can improve multimodal alignment [40], while stylometric and semantic similarity features support joint examination of linguistic style, semantic meaning, and visual-textual correspondence [41]. Overall, the literature shows that multimodal fake news detection has progressed from simple text-image fusion toward uncertainty-aware, explainable, adaptive, domain-sensitive, and inconsistency-oriented frameworks. However, existing studies still provide limited integration of modality-specific evidence, cross-modal inconsistency, confidence estimation, expert disagreement, and adaptive fusion within one decision mechanism. This gap motivates the proposed T-CAMIF framework, which combines cross-modal inconsistency modelling with trust-aware multimodal fusion for text-image fake news detection in social media.

3. Methodology

3.1 Research Design

This study adopted a quantitative experimental research design to develop and evaluate a trust-aware multimodal fake-news detection framework for paired text-image social media posts. The task was formulated as a supervised binary classification problem, where each instance consisted of textual content, an associated image, and a binary class label representing fake/rumor or real/non-rumor content. The proposed framework, named **T-CAMIF: Trust-Aware Cross-Modal Inconsistency Fusion**, was designed to classify paired text-image posts by jointly modelling textual semantics, visual representations, and cross-modal inconsistency.

The experimental pipeline followed a benchmark evaluation structure consisting of dataset preparation, preprocessing, train-validation-test partitioning, text and image feature extraction, cross-modal inconsistency construction, evidence-learner training, trust-aware fusion, and final fake/real prediction. The proposed model was compared with unimodal and multimodal baseline models, and its performance was further examined through ablation analysis, robustness testing, cross-platform evaluation, multi-seed validation, and paired bootstrap comparison. The complete workflow is illustrated in Fig. 1.

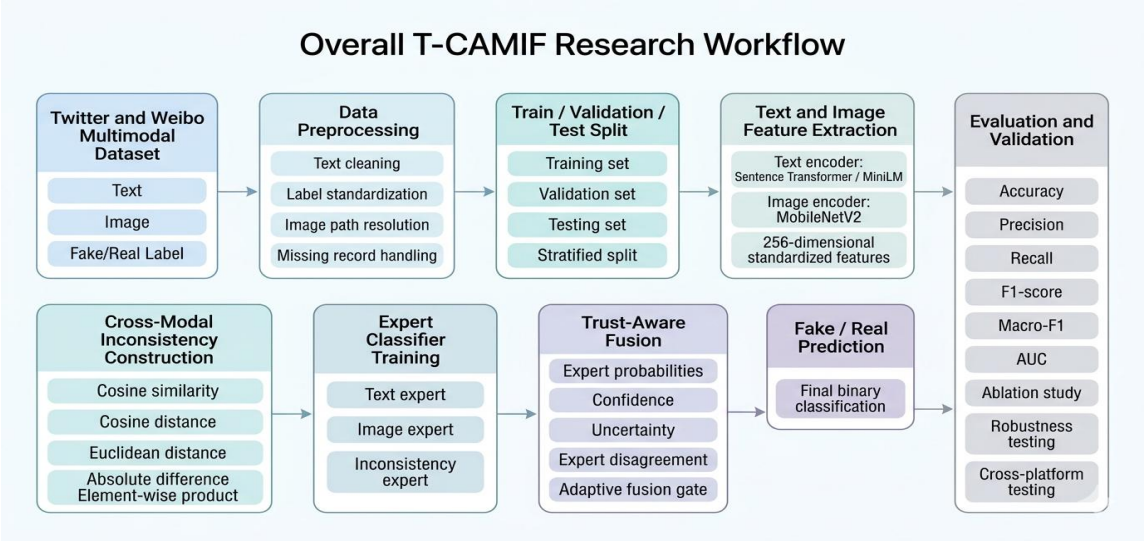

Fig. 1. Overall T-CAMIF research workflow.

3.2 Dataset Description and Data Source

This study used a publicly available Twitter-Weibo multimodal fake-news dataset hosted on Figshare [42]. The dataset contains social media records from Twitter and Weibo, where each instance includes textual content, an associated image, and a fake/real label. The dataset was selected because the proposed T-CAMIF framework requires paired text-image samples to examine whether visual evidence supports, contradicts, or weakly aligns with the textual claim.

Each record was organized into three components: text, image, and label. The text represented the social media post or claim, the image represented the accompanying visual evidence, and the label indicated whether the sample was fake/rumor or real/non-rumor. Labels were standardized into binary form, where **0** represented real/non-rumor content and **1** represented fake/rumor content. Image paths were resolved before feature extraction to ensure correct text-image pairing. Because the dataset is publicly available for research use, this study did not require new data scraping, direct user interaction, or primary data collection. A separate data availability statement is provided at the end of the manuscript to support transparency and reproducibility.

3.3 Dataset Partitioning and Class Distribution

The accessible dataset consisted of paired text-image records from Twitter and Weibo. Since the study required multimodal posts with fake/real labels, the dataset was treated as a benchmark dataset for supervised multimodal fake-news classification.

The dataset was divided into training, validation, and testing subsets using **stratified partitioning** to preserve the fake/real class distribution across subsets. The training set was used to fit the evidence learners and the final trust-aware fusion gate. The validation set was used for model selection and threshold tuning, while the test set was reserved for final evaluation. The final split ratio was 68:12:20 for the Twitter, Weibo, and combined datasets.

Table 1. Dataset partitioning after preprocessing

Dataset	Training	Validation	Testing	Split ratio
Twitter	11,111	1,961	3,268	68:12:20
Weibo	5,721	1,010	1,683	68:12:20
Combined	16,832	2,971	4,951	68:12:20

Table 2. Class distribution in training and testing sets

Dataset	Train fake	Train real	Test fake	Test real
Twitter	5,600	5,511	1,582	1,686
Weibo	2,847	2,874	857	826
Combined	8,447	8,385	2,439	2,512

The class distribution shows that both the platform-specific and combined datasets were approximately balanced. This supports the reliable use of accuracy, F1-score, macro-F1, and AUC for benchmark evaluation.

3.4 Data Preprocessing

The preprocessing phase transformed raw multimodal records into a consistent modelling format suitable for supervised classification. Text preprocessing involved removing or normalizing URLs, user mentions, excessive whitespace, and noisy social media artifacts. Labels from the Twitter and Weibo subsets were converted into a unified binary scheme, where 0 represented real/non-rumor content and 1 represented fake/rumor content.

Image paths were checked and resolved before feature extraction to ensure that each textual record was correctly linked to its corresponding image. Records with unavailable, unreadable, or unusable image files were excluded from the modelling pipeline. The Twitter subset retained complete image availability, while the Weibo subset retained approximately 87.8%–87.9% image availability. Overall, the combined dataset retained approximately 95.9% image availability after preprocessing.

After text and image preprocessing, all extracted feature representations were standardized before classifier training. Standardization was applied to improve numerical stability and to ensure that textual, visual, and cross-modal inconsistency features were comparable during model learning. This preprocessing strategy produced a clean and consistent benchmark dataset for evaluating the proposed T-CAMIF framework.

3.5 Feature Extraction and Cross-Modal Representation

3.5.1 Text Feature Extraction

Textual features were extracted using a Sentence Transformer/MiniLM-based encoder. Each cleaned social media post was converted into a dense semantic embedding that captured contextual meaning from short and informal textual content. The extracted textual embeddings were reduced to 256 dimensions and standardized before classifier training. Dimensionality reduction produced a compact representation, while standardization improved numerical stability and made the text features comparable with the visual feature space.

3.5.2 Image Feature Extraction

Visual features were extracted from the associated images using a MobileNetV2-based encoder. MobileNetV2 was selected because it provides computationally efficient visual representation learning while maintaining suitability for large-scale image feature extraction. The extracted visual embeddings were reduced to 256 dimensions and standardized before model training. This produced a compact image representation aligned with the dimensionality of the textual embeddings.

3.5.3 Cross-Modal Inconsistency Construction

The core feature-engineering component of T-CAMIF was the construction of cross-modal inconsistency features. Let the standardized text feature vector be denoted as:

$$\mathrm{x}_t \in \mathbb{R}^{256} \qquad (1)$$

and the standardized image feature vector be denoted as:

$$\mathrm{x}_v \in \mathbb{R}^{256} \qquad (2)$$

Since both vectors were represented in the same dimensional space, their relationship was estimated using similarity, distance, deviation, and interaction-based measures. The cosine similarity between the text and image vectors was computed as:

$$S_{\cos}(\mathrm{x}_t, \mathrm{x}_v) = \frac{\mathrm{x}_t \cdot \mathrm{x}_v}{\| \mathrm{x}_t \| \ \| \mathrm{x}_v \|} \qquad (3)$$

where $\mathbf{x}_t \cdot \mathbf{x}_v$denotes the dot product between the text and image vectors, and $\| \mathbf{x}_t \|$and $\| \mathbf{x}_v \|$denote their vector norms. Cosine distance was then calculated as:

$$D_{\cos} = 1 - S_{\cos}(\mathrm{x}_t, \mathrm{x}_v) \qquad (4)$$

The Euclidean distance between the two modality vectors was computed as:

$$D_{\mathrm{euc}} = \| \mathbf{x}_t - \mathbf{x}_v \|_2 \qquad (5)$$

Feature-wise deviation was captured through the absolute difference vector:

$$\mathrm{D}_{\mathrm{abs}} = | \mathrm{x}_t - \mathrm{x}_v | \qquad (6)$$

The element-wise interaction between both modalities was computed as:

$$\mathrm{P}_{\mathrm{elem}} = \mathrm{x}_t \odot \mathrm{x}_v \qquad (7)$$

where $\odot$denotes element-wise multiplication. The final cross-modal inconsistency representation was constructed by concatenating the similarity, distance, deviation, and interaction features:

$$\mathrm{z}_{\mathrm{inc}} = [S_{\cos}; D_{\cos}; D_{\mathrm{euc}}; \mathrm{D}_{\mathrm{abs}}; \mathrm{P}_{\mathrm{elem}}] \qquad (8)$$

where $\mathbf{z}_{\mathrm{inc}}$represents the final inconsistency vector used as an independent evidence channel in the proposed framework. This representation captures semantic closeness, cross-modal mismatch, vector-level separation, feature-wise deviation, and modality interaction. The inconsistency representation was treated as a separate evidence source because misinformation may arise not only from textual or visual content independently, but also from a misleading relationship between the two modalities. The overall architecture is shown in Fig. 2.

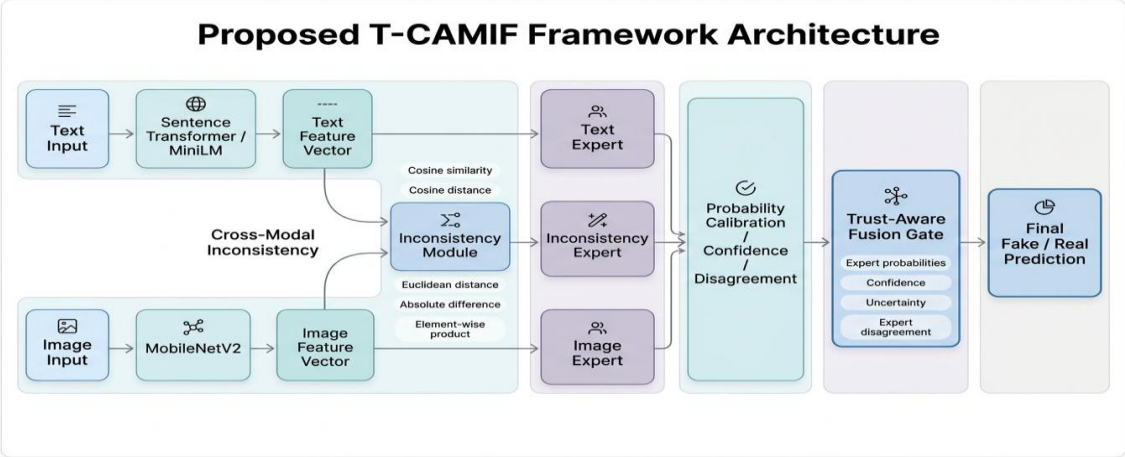

Fig. 2. Proposed T-CAMIF framework architecture.

3.6 Model Development and Trust-Aware Fusion

3.6.1 Evidence Learner Design

The modelling pipeline used five evidence learners to capture modality-specific, cross-modal, and fused multimodal evidence. Three learners were used as primary expert classifiers: a text expert, an image expert, and an inconsistency expert. The text expert was trained on MiniLM-based semantic embeddings, the image expert was trained on MobileNetV2 visual embeddings, and the inconsistency expert was trained on cross-modal inconsistency features. Two additional multimodal learners were also included: a simple concatenation fusion model and a static full-fusion model. The simple concatenation model combined text and image features directly, whereas the static full-fusion model combined text, image, and inconsistency features without adaptive trust weighting.

All five evidence learners were implemented using logistic regression with balanced class weighting. Logistic regression was selected to provide a transparent and controlled evaluation of the proposed evidence representations. Using the same classifier across all evidence learners ensured that performance differences were mainly due to feature representation and fusion strategy rather than classifier complexity. The classifier used the **lbfgs** solver, a maximum of 3000 iterations, and a fixed random seed to support reproducibility. Balanced class weighting was applied to reduce bias toward either fake or real classes. Each evidence learner generated a fake-class probability, and these probability outputs were used as inputs to the trust-aware fusion gate.

3.6.2 Baseline Model Construction

The baseline models were designed to examine whether the proposed T-CAMIF framework improved over unimodal, inconsistency-only, and conventional multimodal fusion approaches. The text-only model used only textual embeddings, the image-only model used only visual embeddings, and the inconsistency-only model used only cross-modal inconsistency features. The simple concatenation fusion model combined text and image features into a single representation, while the static full-fusion model combined text, image, and inconsistency features without adaptive trust estimation.

These baselines provided a structured comparison among single-modality learning, inconsistency-based learning, conventional multimodal fusion, and the proposed trust-aware fusion mechanism. The same logistic regression configuration was used across all baseline models to ensure a fair comparison and to isolate the contribution of the proposed feature representations and fusion strategy.

3.6.3 Trust-Aware Fusion Gate

The final T-CAMIF model used a learned trust-aware fusion gate. The fusion gate was implemented as a logistic regression meta-classifier trained on evidence-level probability features derived from the five evidence learners. Let the fake-class probability outputs of the five evidence learners be represented as:

$$\mathrm{P} = [p_t, p_v, p_i, p_c, p_f] \qquad (9)$$

where p_t, p_v, p_i, p_c, and p_f denote the probabilities produced by the text expert, image expert, inconsistency expert, simple concatenation model, and static full-fusion model, respectively.

The confidence of each evidence learner was estimated from its distance from the uncertainty point of 0.5:

$$c_k = | p_k - 0.5 | \qquad (10)$$

where c_k denotes the confidence score of the k-th evidence learner. A value close to 0 indicates higher uncertainty, while a value closer to 0.5 indicates stronger confidence. The confidence vector was defined as:

$$\mathbf{C} = [c_t, c_v, c_i, c_c, c_f] \qquad (11)$$

To capture agreement and disagreement among the evidence learners, three aggregate probability statistics were computed: mean probability, standard deviation, and probability range:

$$\mu_P = \frac{1}{5}\sum_{k=1}^{5} p_k \qquad (12)$$

$$\sigma_P = \sqrt{\frac{1}{5}\sum_{k=1}^{5}(p_k - \mu_P)^2} \qquad (13)$$

$$r_P = \max(\mathbf{P}) - \min(\mathbf{P}) \qquad (14)$$

Here, σ_P and r_P represent disagreement indicators. Higher values indicate greater divergence among the evidence learners.

The trust-gate input vector was constructed by concatenating the evidence probabilities, confidence scores, and aggregate agreement-disagreement indicators:

$$\mathrm{z} = [\mathrm{P}, \mathrm{C}, \mu_P, \sigma_P, r_P] \qquad (15)$$

The final fake-class probability was produced by the trust-aware fusion function:

$$\hat{y}_{\mathrm{prob}} = g(\mathbf{z}) \qquad (16)$$

where $g(\cdot)$ denotes the logistic regression meta-classifier used as the trust-aware fusion gate. The final class label was assigned using a validation-tuned threshold τ:

$$\hat{y} = \begin{cases} 1, & \hat{y}_{\mathrm{prob}} \geq \tau \\ 0, & \hat{y}_{\mathrm{prob}} < \tau \end{cases} \qquad (17)$$

where 1 represents fake/rumor content and 0 represents real/non-rumor content. This design allowed the final prediction to consider not only the individual evidence probabilities, but also the confidence and disagreement patterns among the evidence learners.

3.7 Evaluation and Validation Protocol

Model performance was evaluated using accuracy, precision, recall, F1-score, macro-F1, and area under the receiver operating characteristic curve. These metrics were selected because the task was formulated as a binary classification problem and the dataset contained approximately balanced fake and real classes. The validation set was used for threshold tuning. The decision threshold was selected by maximizing validation F1-score, while the test set was reserved only for final performance evaluation. This ensured that model selection and final testing were conducted on separate data partitions.

The evaluation protocol included baseline comparison, ablation analysis, robustness testing, cross-platform evaluation, multi-seed validation, and paired bootstrap comparison. Baseline comparison assessed the proposed model against unimodal, inconsistency-only, and conventional multimodal fusion models. Ablation analysis evaluated the contribution of each model component. Robustness testing examined model sensitivity under missing, masked, noisy, or mismatched modality conditions. Cross-platform evaluation assessed transfer performance between Twitter and Weibo. Multi-seed validation examined computational reproducibility, while paired bootstrap comparison assessed the statistical reliability of performance differences. Overall, the evaluation protocol assessed

classification performance, component contribution, robustness to modality perturbation, cross-platform transferability, computational reproducibility, and statistical reliability.

3.8 Ethical Considerations

This study used an existing publicly available benchmark dataset and did not involve direct user interaction, new data scraping, or primary data collection. No user profiling or personal identification was performed. The fake/real labels were treated as dataset annotations for computational modelling rather than as judgments about specific individuals, groups, or communities.

Because fake-news detection may influence content moderation, public communication, and information verification practices, the study emphasized interpretable evidence modelling through evidence-level outputs, cross-modal inconsistency features, uncertainty indicators, confidence scores, and trust-aware fusion. The proposed framework is intended to support research on multimodal misinformation detection and should not be used as a fully automated moderation system without human review.

4. Results

4.1 Overall Classification Performance

The proposed T-CAMIF framework was evaluated on the combined Twitter-Weibo dataset using accuracy, precision, recall, F1-score, macro-F1, and AUC. As reported in Table 3, T-CAMIF achieved the strongest in-domain performance among the evaluated models, with an accuracy of 0.9061, precision of 0.8658, recall of 0.9578, F1-score of 0.9095, macro-F1 of 0.9059, and AUC of 0.9709.

Table 3. Main performance comparison on the combined Twitter-Weibo dataset

Model	Accuracy	Precision	Recall	F1-score	Macro-F1	AUC
Text-only	0.7703	0.7151	0.8872	0.7919	0.7678	0.8594
Image-only	0.8065	0.7347	0.9504	0.8287	0.8032	0.9245
Inconsistency-only	0.7782	0.7441	0.8380	0.7883	0.7777	0.8773
Simple concat fusion	0.9016	0.8620	0.9528	0.9052	0.9015	0.9695
Static full fusion	0.9004	0.8580	0.9561	0.9044	0.9003	0.9689
T-CAMIF trust gate	0.9061	0.8658	0.9578	0.9095	0.9059	0.9709

The results show that multimodal models outperformed unimodal models, confirming the complementary value of textual and visual evidence. The image-only model performed better than the text-only model, indicating that visual features were highly informative in this benchmark setting. The inconsistency-only model also achieved competitive performance relative to the text-only model, suggesting that text-image mismatch contains useful classification signals.

Overall, T-CAMIF achieved the best in-domain results across all reported metrics. The improvement over unimodal and inconsistency-only baselines was substantial, while the improvement over simple concatenation and static full fusion was smaller but consistent. This indicates that adaptive trust-aware fusion provided additional value beyond conventional multimodal feature combination, although the margin over the strongest fusion baseline should be interpreted as incremental rather than large.

4.2 Fake and Real Prediction Analysis

The final fake/real prediction behaviour of T-CAMIF was examined using the confusion matrix in Fig. 3. On the combined Twitter-Weibo test set, the model correctly classified **2,150 real samples as real** and **2,336 fake samples as fake**. It misclassified **362 real samples as fake** and **103 fake samples as real**. The lower number of

false negatives indicates that the model was effective in identifying fake content, which is important because undetected misinformation may continue to circulate.

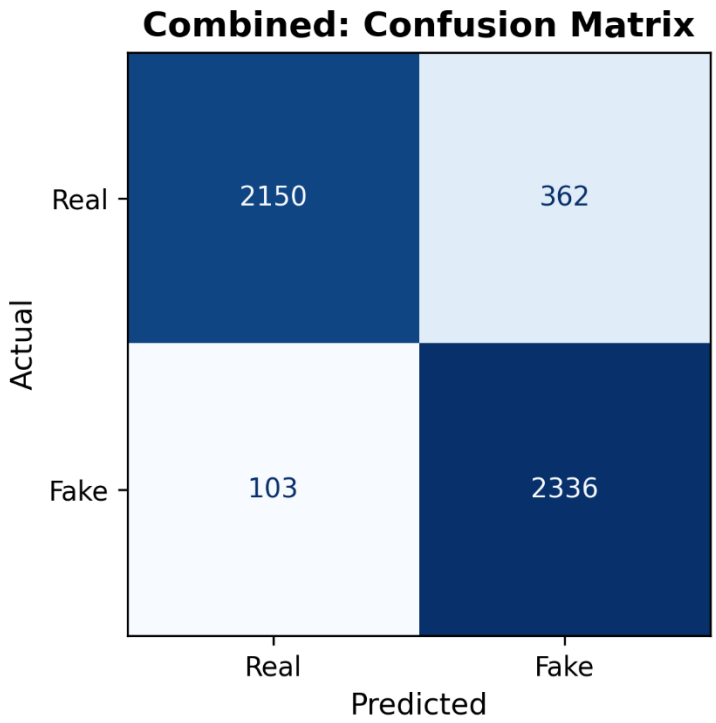

Fig. 3. Confusion matrix of the proposed T-CAMIF model on the combined Twitter-Weibo test set.

Class-wise results in Table 4 further confirm the balanced predictive performance of the model. On the combined dataset, T-CAMIF achieved an F1-score of **0.9105** for the real class and **0.9114** for the fake class. The fake-class recall of **0.9299** shows that most fake samples were correctly detected, while the real-class precision of **0.9291** indicates reliable identification of real content.

Table 4. Class-wise fake and real prediction performance of T-CAMIF

Dataset	Class	Precision	Recall	F1-score	Support
Combined	Real	0.9291	0.8925	0.9105	2,512
Combined	Fake	0.8936	0.9299	0.9114	2,439
Twitter	Real	0.9976	0.9947	0.9961	1,686
Twitter	Fake	0.9943	0.9975	0.9959	1,582
Weibo	Real	0.8941	0.8075	0.8486	826
Weibo	Fake	0.8303	0.9078	0.8673	857

Dataset-specific results show that the model achieved near-perfect class-wise performance on the Twitter subset, whereas the Weibo subset was comparatively more challenging, particularly for real-class recall. This suggests that platform-specific differences in language, event distribution, visual style, and posting behaviour affected prediction stability.

4.3 ROC-Based Discrimination

The ROC curve in Fig. 4 shows strong probability-level discrimination between fake and real classes. The curve remains well above the diagonal reference line, indicating effective class separation across different decision

thresholds. The AUC of **0.9709** further confirms that T-CAMIF achieved strong separability and reliable probabilistic discrimination on the combined Twitter-Weibo test set.

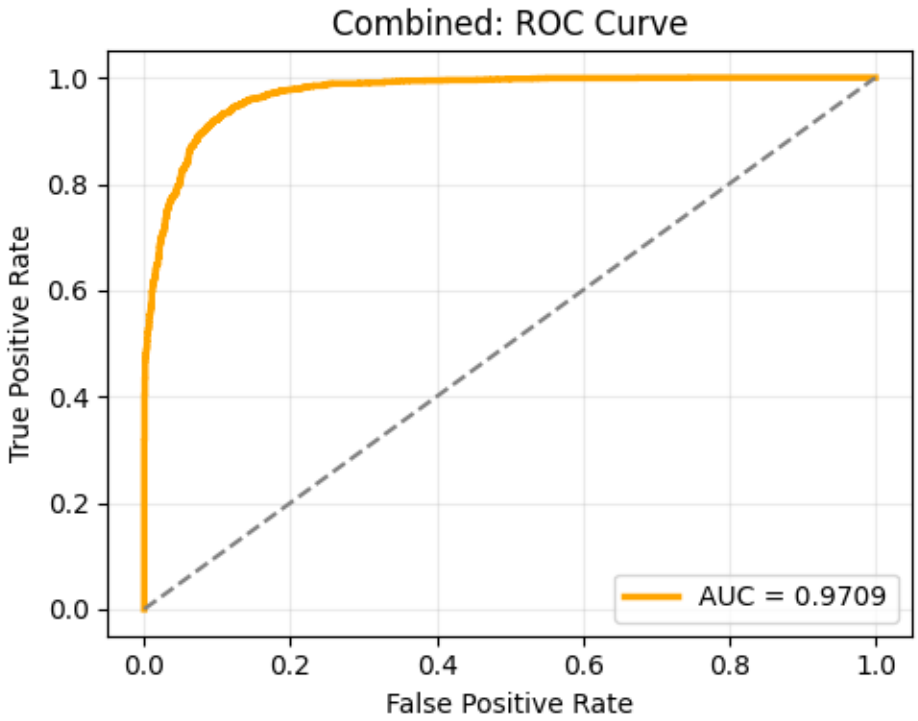

Fig. 4. ROC curve of the proposed T-CAMIF model on the combined Twitter-Weibo test set.

4.4 Ablation Study

The ablation study assessed the contribution of each component in the proposed framework. As shown in **Fig. 5**, the complete T-CAMIF model achieved the highest overall performance, with an accuracy of **0.9061**, F1-score of **0.9095**, macro-F1 of **0.9059**, and AUC of **0.9709**. The figure compares the F1-score and AUC of different ablation variants, highlighting the effect of removing or isolating specific components.

The results show that the best performance was obtained when text, image, cross-modal inconsistency, and trust-aware fusion were combined. The **without image branch** and **without text branch** variants produced lower scores, confirming that both modalities contributed to the final prediction. The **inconsistency-only** variant also achieved meaningful performance, showing that text-image mismatch provides useful classification evidence. However, the strongest results were achieved only when inconsistency features were combined with textual, visual, and trust-aware fusion components.

The trust-aware gate improved over static full fusion, indicating that confidence and disagreement indicators contributed to the final prediction. However, the improvement over the simple text-image concatenation model, represented by the **without inconsistency module** variant, was modest. This suggests that the trust-aware mechanism provides an incremental but consistent performance gain over strong conventional fusion.

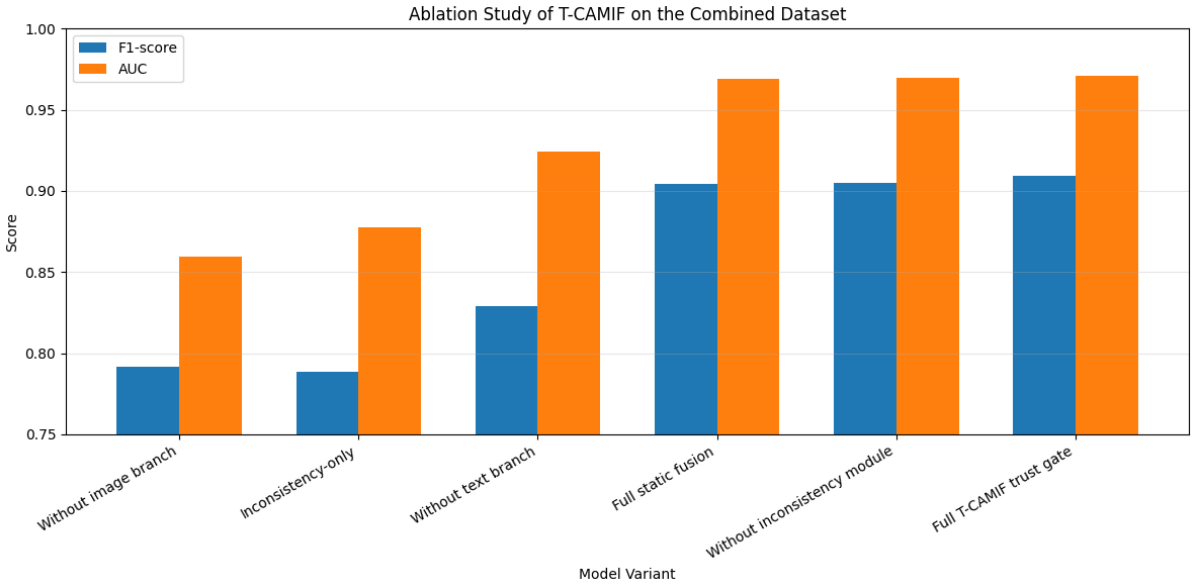

Fig. 5. Ablation study of T-CAMIF on the combined Twitter-Weibo dataset.

4.5 Robustness Testing

Robustness testing was performed to examine the sensitivity of T-CAMIF under different modality perturbation conditions, including clean test data, missing text, missing image, noisy image features, masked text features, and mismatched image-text pairs. As illustrated in Fig 6, the clean test condition reproduced the main evaluation performance, with an F1-score of 0.9095 and an AUC of 0.9709.

Under partial perturbation, the model remained relatively stable. In particular, noisy image features and masked text features resulted in F1-scores of 0.8788 and 0.8793, respectively, showing that T-CAMIF can tolerate moderate degradation in one modality. By contrast, performance declined when an entire modality was unavailable, especially in the missing-image condition, where the F1-score decreased to 0.7036.

The most substantial reduction in performance occurred under mismatched image-text pairs, with an F1-score of 0.6050 and an AUC of 0.6739. This result highlights the importance of correct semantic correspondence between textual and visual content in multimodal fake-news detection. Overall, the robustness analysis shows that T-CAMIF remains effective under moderate perturbation but is more vulnerable to severe modality loss and disrupted cross-modal alignment.

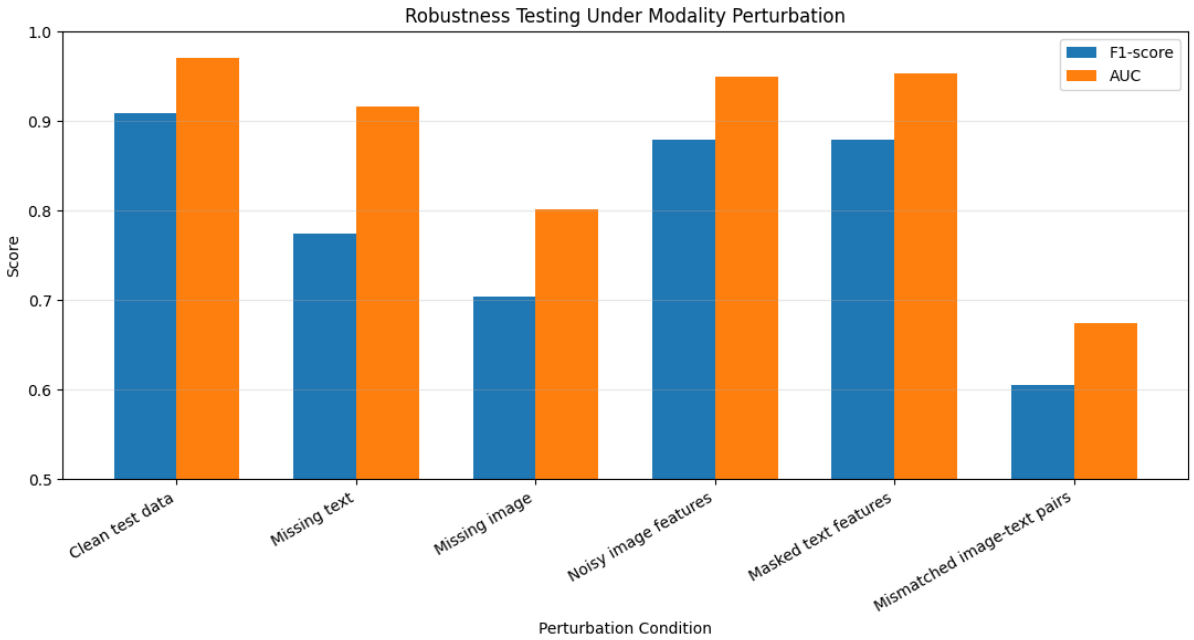

Fig. 6. Robustness performance of T-CAMIF under modality perturbation conditions.

4.6 Cross-Platform Generalization

Cross-platform testing was conducted to evaluate whether models trained on one platform could transfer effectively to another. As shown in Table 5, transfer performance was lower than in-domain performance for both Twitter-to-Weibo and Weibo-to-Twitter settings. In the Twitter-to-Weibo setting, the text-only model achieved the highest accuracy of **0.5330**, while T-CAMIF achieved an accuracy of **0.4949**. In the Weibo-to-Twitter setting, static full fusion achieved the highest accuracy of **0.5245**, while T-CAMIF achieved an accuracy of **0.4893**.

Table 5. Cross-platform generalization results

Transfer setting	Model	Accuracy	F1-score	AUC
Twitter-to-Weibo	Text-only	0.5330	0.6348	0.5244
Twitter-to-Weibo	T-CAMIF trust gate	0.4949	0.6272	0.4818
Weibo-to-Twitter	Static full fusion	0.5245	0.5798	0.5276
Weibo-to-Twitter	T-CAMIF trust gate	0.4893	0.5439	0.4882

These results show that transfer between Twitter and Weibo remained challenging. The lower cross-platform scores suggest that language, event distribution, visual style, and platform-specific posting behaviour influenced multimodal fake-news detection. Therefore, cross-platform transferability should be treated as a limitation of the current framework and an important direction for future work.

4.7 Statistical Stability and Comparative Validation

Additional validation was conducted to assess the statistical stability and comparative reliability of the proposed T-CAMIF framework. As shown in Table 6, the multi-seed validation produced identical scores across repeated runs, with accuracy of **0.9061 ± 0.0000**, F1-score of **0.9095 ± 0.0000**, macro-F1 of **0.9059 ± 0.0000**, and AUC of **0.9709 ± 0.0000**. This result indicates computational reproducibility under the fixed extracted-feature setting.

The paired bootstrap comparison showed that, T-CAMIF achieved statistically significant F1-score improvements over the inconsistency-only, text-only, image-only, and static full-fusion baselines. The improvement over simple concatenation was smaller and close to the conventional significance threshold, indicating that the trust-

aware gate provides additional benefit but should be interpreted as an incremental improvement over the strongest conventional fusion baseline.

Table 6. Summary of statistical stability and comparative validation

Validation check	Main result
Multi-seed validation	Accuracy = 0.9061 ± 0.0000; F1-score = 0.9095 ± 0.0000; Macro-F1 = 0.9059 ± 0.0000; AUC = 0.9709 ± 0.0000
Bootstrap: T-CAMIF vs inconsistency-only	ΔF1 = 0.1210; 95% CI [0.1096, 0.1324]; $p < 0.001$
Bootstrap: T-CAMIF vs text-only	ΔF1 = 0.1177; 95% CI [0.1073, 0.1278]; $p < 0.001$
Bootstrap: T-CAMIF vs image-only	ΔF1 = 0.0807; 95% CI [0.0715, 0.0901]; $p < 0.001$
Bootstrap: T-CAMIF vs static full fusion	ΔF1 = 0.0051; 95% CI [0.0012, 0.0093]; $p = 0.008$
Bootstrap: T-CAMIF vs simple concat fusion	ΔF1 = 0.0043; 95% CI [-0.0000, 0.0086]; $p = 0.052$

Overall, the results confirm that T-CAMIF was reproducible under the fixed experimental setting and achieved statistically supported improvements over most baseline configurations. The comparison with simple concatenation suggests that the trust-aware fusion mechanism provides a modest but meaningful enhancement over conventional multimodal fusion.

5. Discussion

The results show that T-CAMIF improved in-domain multimodal fake-news detection by integrating textual evidence, visual evidence, and cross-modal inconsistency through a trust-aware fusion gate. On the combined Twitter-Weibo dataset, T-CAMIF achieved the strongest performance, with accuracy of 0.9061, F1-score of 0.9095, macro-F1 of 0.9059, and AUC of 0.9709. The improvement over unimodal and inconsistency-only baselines was substantial, while the gain over simple concatenation and static full fusion was smaller but consistent. Therefore, the trust-aware mechanism should be interpreted as an incremental but meaningful improvement over strong conventional fusion baselines.

This improvement can be explained by the evidence-learner structure of T-CAMIF. Instead of merging all features at the initial stage, the framework preserves separate text, image, and cross-modal inconsistency streams before final prediction. This enables the model to retain modality-specific information while also capturing whether textual and visual evidence are semantically aligned, weakly related, or contradictory. The inconsistency-only model achieved meaningful performance, confirming that text-image mismatch contains useful classification evidence. However, the strongest results were obtained when inconsistency features were combined with textual, visual, and trust-aware fusion components, showing that cross-modal mismatch is most effective as complementary evidence.

The trust-aware fusion gate further improved decision-making by incorporating expert probabilities, confidence scores, uncertainty indicators, and disagreement measures. This allowed the model to consider not only what each learner predicted, but also how reliable or divergent the evidence streams were. This is important because social media posts often contain noisy captions, incomplete claims, irrelevant images, or conflicting multimodal evidence. Robustness results support this interpretation: T-CAMIF remained comparatively stable under noisy image features and masked text features, but performance declined when an entire modality was unavailable or when text-image correspondence was disrupted. Thus, reliable multimodal fake-news detection depends on correct semantic alignment between text and image.

Compared with existing literature, the findings are consistent with dynamic fusion research, which shows that modality contributions should be adjusted according to evidence reliability [1]. The uncertainty-sensitive design of

T-CAMIF is also aligned with evidential and Bayesian uncertainty research, which emphasizes confidence estimation under uncertain or unreliable conditions [36], [37]. The contribution of the text branch is consistent with attention-based and bidirectional language representation research, where contextual modelling supports semantic understanding of short and informal text [38], [39]. Similarly, the role of cross-modal inconsistency supports hierarchical cross-modal interaction research, where structured text-image relationships are treated as important signals for multimodal fake-news detection [43].

Cross-platform evaluation showed that transfer between Twitter and Weibo remained challenging. T-CAMIF performed strongly in the combined in-domain setting but achieved weaker results when trained on one platform and tested on another. This suggests that language, event distribution, visual style, cultural context, and platform-specific posting behaviour influence multimodal fake-news detection. The results were obtained under the selected benchmark setting, and cross-platform evaluation showed that transfer across social media platforms remains challenging. Future work should evaluate the framework using larger multilingual datasets, event-aware partitioning, stronger vision-language encoders, and domain-adaptive learning.

6. Conclusion

This study proposed **T-CAMIF**, a trust-aware cross-modal inconsistency fusion framework for multimodal fake-news detection in paired text-image social media posts. The framework addresses the limitation of conventional multimodal models that directly merge textual and visual features without explicitly assessing their reliability or semantic consistency. By combining textual evidence, visual evidence, and cross-modal inconsistency with confidence and disagreement indicators, T-CAMIF provides an adaptive mechanism for fake/real classification.

Experimental results showed that T-CAMIF achieved strong in-domain performance on the combined Twitter-Weibo dataset and outperformed unimodal, inconsistency-only, simple concatenation, and static full-fusion baselines. The ablation analysis confirmed the contribution of cross-modal inconsistency and trust-aware fusion, while robustness testing showed that the model remained comparatively stable under noisy image and masked text conditions. The results also showed that reliable text-image correspondence is important for effective multimodal fake-news detection.

This study contributes to trustworthy media analytics by demonstrating that cross-modal inconsistency can serve as an informative and interpretable evidence source. It also highlights the value of confidence, uncertainty, and expert disagreement in multimodal decision-making. Cross-platform evaluation showed that platform variation remains a key challenge. Future work should explore stronger vision-language encoders, larger multilingual datasets, event-aware partitioning, domain adaptation, and advanced uncertainty-aware fusion to improve robustness and generalization across platforms.

References

1. H. Lv, W. Yang, Y. Yin, F. Wei, J. Peng, and H. Geng, “MDF-FND: A dynamic fusion model for multimodal fake news detection,” Knowledge-Based Systems, vol. 317, Art. no. 113417, 2025.
2. Z. Li, J. Yang, X. Wang, J. Lei, S. Li, and J. Zhang, “Uncertainty-aware disentangled representation learning for multimodal fake news detection,” Information Processing & Management, vol. 62, no. 5, Art. no. 104190, 2025.
3. J. Qiao, X. Li, C. Gao, L. Wu, J. Feng, and Z. Wang, “Improving multimodal fake news detection by leveraging cross-modal content correlation,” Information Processing & Management, vol. 62, no. 5, Art. no. 104120, 2025.
4. S. Wei, Z. Wang, M. Li, X. Liu, and B. Wu, “DCCMA-Net: Disentanglement-based cross-modal clues mining and aggregation network for explainable multimodal fake news detection,” Information Processing & Management, vol. 62, no. 4, Art. no. 104089, 2025.
5. T. M. H. Gedara, V. Loia, and S. Tomasiello, “A fuzzy-based multimodal approach for interpretable fake news detection,” Applied Soft Computing, vol. 179, Art. no. 113277, 2025.
6. J. Jing, H. Wu, J. Sun, X. Fang, and H. Zhang, “Multimodal fake news detection via progressive fusion networks,” Information Processing & Management, vol. 60, no. 1, Art. no. 103120, 2023.
7. J. Xue, Y. Wang, Y. Tian, Y. Li, L. Shi, and L. Wei, “Detecting fake news by exploring the consistency of multimodal data,” Information Processing & Management, vol. 58, no. 5, Art. no. 102610, 2021.
8. B. Wang, X. Li, C. Li, S. Wang, and W. Gao, “Escaping the neutralization effect of modality features fusion in multimodal fake news detection,” Information Fusion, vol. 111, Art. no. 102500, 2024.
9. Q. Li, M. Gao, G. Zhang, W. Zhai, J. Chen, and G. Jeon, “Towards multimodal disinformation detection by vision-language knowledge interaction,” Information Fusion, vol. 102, Art. no. 102037, 2024.
10. F. Wu, S. Chen, G. Gao, Y. Ji, and X. Y. Jing, “Balanced multi-modal learning with hierarchical fusion for fake news detection,” Pattern Recognition, vol. 164, Art. no. 111485, 2025.

11. W. Cui, X. Zhang, and M. Shang, “Multi-view mutual learning network for multimodal fake news detection,” Expert Systems with Applications, vol. 279, Art. no. 127407, 2025.
12. Y. Zhou, Y. Yang, Q. Ying, Z. Qian, and X. Zhang, “Multi-modal fake news detection on social media via multi-grained information fusion,” in Proc. 2023 ACM Int. Conf. Multimedia Retrieval, Jun. 2023, pp. 343–352.
13. Y. Chen, D. Li, P. Zhang, J. Sui, Q. Lv, L. Tun, and L. Shang, “Cross-modal ambiguity learning for multimodal fake news detection,” in Proc. ACM Web Conf. 2022, Apr. 2022, pp. 2897–2905.
14. Y. Wang et al., “EANN: Event adversarial neural networks for multi-modal fake news detection,” in Proc. 24th ACM SIGKDD Int. Conf. Knowledge Discovery & Data Mining, Jul. 2018, pp. 849–857.
15. D. Khattar, J. S. Goud, M. Gupta, and V. Varma, “MVAE: Multimodal variational autoencoder for fake news detection,” in Proc. World Wide Web Conf., May 2019, pp. 2915–2921.
16. S. Singhal, R. R. Shah, T. Chakraborty, P. Kumaraguru, and S. I. Satoh, “SpotFake: A multi-modal framework for fake news detection,” in Proc. 2019 IEEE 5th Int. Conf. Multimedia Big Data (BigMM), Sep. 2019, pp. 39–47.
17. S. Singhal, A. Kabra, M. Sharma, R. R. Shah, T. Chakraborty, and P. Kumaraguru, “SpotFake+: A multimodal framework for fake news detection via transfer learning,” in Proc. AAAI Conf. Artificial Intelligence, vol. 34, no. 10, Apr. 2020, pp. 13915–13916.
18. Z. Ma, M. Luo, H. Guo, Z. Zeng, Y. Hao, and X. Zhao, “Event-Radar: Event-driven multi-view learning for multimodal fake news detection,” in Proc. 62nd Annu. Meeting Association for Computational Linguistics, vol. 1, Aug. 2024, pp. 5809–5821.
19. X. Liu et al., “FKA-Owl: Advancing multimodal fake news detection through knowledge-augmented LVLMs,” in Proc. 32nd ACM Int. Conf. Multimedia, Oct. 2024, pp. 10154–10163.
20. B. Wang, S. Wang, C. Li, R. Guan, and X. Li, “Harmfully manipulated images matter in multimodal misinformation detection,” in Proc. 32nd ACM Int. Conf. Multimedia, Oct. 2024, pp. 2262–2271.
21. Q. Ying, X. Hu, Y. Zhou, Z. Qian, D. Zeng, and S. Ge, “Bootstrapping multi-view representations for fake news detection,” in Proc. AAAI Conf. Artificial Intelligence, vol. 37, no. 4, Jun. 2023, pp. 5384–5392.
22. H. Wang, J. Guo, S. Liu, P. Chen, and X. Li, “SEPM: Multiscale semantic enhancement-progressive multimodal fusion network for fake news detection,” Expert Systems with Applications, vol. 283, Art. no. 127741, 2025.
23. Y. Meng, H. Wang, J. Zhao, Y. Sun, and M. Shao, “Dynamic hierarchical memory improved mixture-of-experts for multimodal fake news detection,” Information Processing & Management, vol. 63, no. 2, Art. no. 104479, 2026.
24. S. Zhang, H. Chen, S. Liu, R. Li, H. Jiang, and L. Dong, “Multimodal misinformation detection based on the multi-granularity consistency,” Neurocomputing, vol. 644, Art. no. 130393, 2025.
25. W. Lu, Y. Tong, and Z. Ye, “DAMMFND: Domain-aware multimodal multi-view fake news detection,” in Proc. AAAI Conf. Artificial Intelligence, vol. 39, no. 1, Apr. 2025, pp. 559–567.
26. C. Comito, L. Caroprese, and E. Zumpano, “Multimodal fake news detection on social media: A survey of deep learning techniques,” Social Network Analysis and Mining, vol. 13, no. 1, Art. no. 101, 2023.
27. M. Nasser et al., “A systematic review of multimodal fake news detection on social media using deep learning models,” Results in Engineering, vol. 26, Art. no. 104752, 2025.
28. S. Abdali, S. Shaham, and B. Krishnamachari, “Multi-modal misinformation detection: Approaches, challenges and opportunities,” ACM Computing Surveys, vol. 57, no. 3, pp. 1–29, 2024.
29. K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, “Fake news detection on social media: A data mining perspective,” ACM SIGKDD Explorations Newsletter, vol. 19, no. 1, pp. 22–36, 2017.
30. X. Zhou and R. Zafarani, “A survey of fake news: Fundamental theories, detection methods, and opportunities,” ACM Computing Surveys, vol. 53, no. 5, pp. 1–40, 2020.
31. D. M. J. Lazer et al., “The science of fake news,” Science, vol. 359, no. 6380, pp. 1094–1096, 2018.
32. X. Zhou, R. Zafarani, K. Shu, and H. Liu, “Fake news: Fundamental theories, detection strategies and challenges,” in Proc. 12th ACM Int. Conf. Web Search and Data Mining, Jan. 2019, pp. 836–837.
33. L. Hu, S. Wei, Z. Zhao, and B. Wu, “Deep learning for fake news detection: A comprehensive survey,” AI Open, vol. 3, pp. 133–155, 2022.
34. A. B. Arrieta et al., “Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI,” Information Fusion, vol. 58, pp. 82–115, 2020.
35. J. Gawlikowski et al., “A survey of uncertainty in deep neural networks,” Artificial Intelligence Review, vol. 56, Suppl. 1, pp. 1513–1589, 2023.
36. M. Sensoy, L. Kaplan, and M. Kandemir, “Evidential deep learning to quantify classification uncertainty,” in Advances in Neural Information Processing Systems, vol. 31, 2018.
37. A. Kendall and Y. Gal, “What uncertainties do we need in Bayesian deep learning for computer vision?” in Advances in Neural Information Processing Systems, vol. 30, 2017.
38. A. Vaswani et al., “Attention is all you need,” in Advances in Neural Information Processing Systems, vol. 30, 2017.
39. J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in Proc. 2019 Conf. North American Chapter Association for Computational Linguistics: Human Language Technologies, vol. 1, Jun. 2019, pp. 4171–4186.
40. A. Radford et al., “Learning transferable visual models from natural language supervision,” in Proc. Int. Conf. Machine Learning, Jul. 2021, pp. 8748–8763.

41. A. Dosovitskiy et al., “An image is worth 16x16 words: Transformers for image recognition at scale,” arXiv preprint arXiv:2010.11929, 2020.
42. A. Ojo, “[dataset] Datasets: fake news multimodal datasets (Twitter and Weibo). Credit: Data 1: Twitter dataset; Data 2: Weibo dataset,” Figshare, version 2, 2025. doi: 10.6084/m9.figshare.28516655.v2.
43. Z. Li, X. Li, J. Yang, X. Wang, S. Li, and J. Zhang, “Hierarchical cross-modal interaction network for multimodal fake news detection,” Neurocomputing, vol. 647, Art. no. 130308, 2025.