DAMMFND: Domain-Aware Multimodal Multi-view Fake News Detection

Weihai Lu[1]*†, Yu Tong[2]†, Zhiqiu Ye[3]

[1]Peking University
[2]Wuhan University
[3]Anhui University
luweihai@pku.edu.cn, yutchina02@gmail.com, r22114015@stu.ahu.edu.cn

Abstract

Recently, multi-domain fake news detection has garnered increasing attention in academia. In particular, the integration of multimodal information into multi-domain fake news detection has emerged as a highly promising research direction. However, this field faces three main challenges: (1) Inaccurate domain identification, where predefined explicit identifiers fail to adapt to the inherent complexity of data; (2) Imbalanced multi-domain data distribution, which may induce negative transfer effects; and (3) Variable multi-domain modal contributions, indicating domain-specific differences in how various modalities influence news veracity assessments. To address these issues, we propose the **D**omain-**A**ware **M**ulti-Modal **M**ulti-View **F**ake **N**ews **D**etection (DAMMFND) framework. DAMMFND effectively extracts more accurate domain information through Domain Disentanglement, while simultaneously mitigating negative transfer between domains. Furthermore, DAMMFND introduces a Domain-Aware Multi-View Discriminator and a Domain-Enhanced Multi-view Decision Layer, which accurately quantify the contribution of domain information to multimodal, multi-view decision-making processes. Extensive experiments conducted on two real-world datasets demonstrate that the proposed model outperforms state-of-the-art baselines.

Code — https://github.com/luweihai/DAMMFND

Introduction

The proliferation of fake news not only severely hinders the healthy development of social media on the internet but also causes actual harm across political, economic, and social dimensions, highlighting the growing importance of automated fake news detection(Zhou et al. 2015; Cui et al. 2019; Zhu et al. 2022a; Li et al. 2024c,b,a). Existing fake news detection methods typically focus on news from a single domain, such as health or politics. However, in the real world, news on social media often spans multiple domains, necessitating existing fake news detection methods to consider multi-domain generalization (Nan et al. 2021; Qi et al. 2019; Zhu et al. 2022b). Additionally, multimodal content (primarily referring to text and image modalities in this paper) acts as a catalyst for fake news propagation (Jing et al. 2023; Wu, Liu, and Zhang 2023; Wei et al. 2022), making multimodal multi-domain fake news detection a promising research direction.

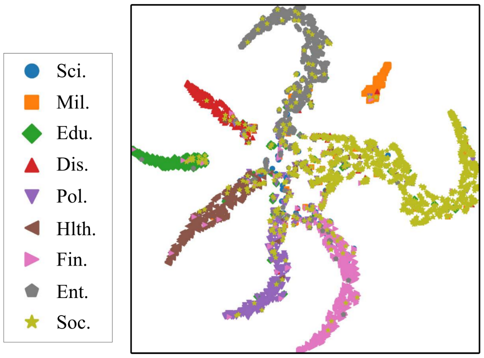


Figure 1: The t-SNE visualization results of the Weibo21 dataset reveal partially overlapping boundaries between different domains, highlighting the inherent complexity of cross-domain classification within this dataset.

KATMF (Song et al. 2021) is the first fake news detection method that combines multi-domain and multimodal approaches, capturing feature distribution differences across news articles from various domains through adversarial multi-task learning and knowledge-enhanced Transformers. MMDFND (Tong et al. 2024) further improves performance in multimodal multi-domain fake news detection by modeling domain commonalities and specificities through an improved PLE module. However, existing multimodal multi-domain fake news detection methods have failed to effectively address the following three challenges:

1. **Inaccurate intra-domain classification**: Current multimodal multi-domain fake news detection models often rely on predefined explicit scenario recognition, which fails to capture the inherent complexity of the data. As illustrated in Figure1, t-SNE visualizations (Van der Maaten and Hinton 2008) of TextCNN-learned (Zhang and Wallace 2015) domain classification representations reveal significant overlap between domains. For instance,

*Corresponding author

†These authors contributed equally.

Domain	Science	Military	Edu.	Disasters	Politics
Real	143	121	243	185	306
Fake	93	222	248	591	546
Total	236	343	491	776	852
Domain	**Health**	**Finance**	**Ent.**	**Society**	**All**
Real	485	959	1,000	1,198	4,640
Fake	515	362	440	1,471	4,488
Total	1,000	1,321	1,440	2,669	9,128

Table 1: Data Statistics of Weibo21

news about technological innovations may simultaneously belong to technology and economic domains, discussing both technical advancements and their market impacts. This ambiguity introduces noise in single-scenario identification, which is amplified by multi-domain modeling modules (e.g., MMOE (Nan et al. 2021) and PLE (Tong et al. 2024)).

2. **Uneven cross-domain data distribution**: Data volume varies significantly across domains, with some domains being data-rich while others are data-sparse. As shown in Table 1, “Society” news far outnumbers ”Science” and “Military” domains, with fake news proportions also varying considerably. This imbalance leads to insufficient learning in data-sparse domains. Moreover, shared parameters may be biased towards data-rich domains, potentially causing negative transfer to data-sparse scenarios where shared information diverges significantly from the target scenario’s features and distribution, thus compromising model performance (Zhang et al. 2022).
3. **Uneven modal contributions across domains**: Domains rely on modal information to varying degrees for decision-making. Some prioritize text authenticity, others focus on image manipulation, while some require joint text-image analysis. For example, an artificially altered image might be classified as fake news in ”political” domains but considered genuine in ”entertainment” contexts. This creates a domain-specific bias in the reliance on textual, visual, and multimodal features. Current multimodal-based methods often overlook this bias during modeling, potentially limiting optimal performance.

To address these challenges, we propose **D**omain-**A**ware **M**ulti-Modal **M**ulti-View **F**ake **N**ews **D**etection (DAMMFND). This approach employs domain disentanglement to extract domain-invariant and domain-specific representations from multimodal news data, improving domain information modeling accuracy and mitigating negative transfer between domains. Additionally, we introduce a Domain-Aware Multi-View Discriminator and decision layer to enhance the multimodal decision process using domain knowledge. Our key contributions are:

1. We propose a Domain-Aware Multi-Modal Multi-View Fake News Detection framework, capable of effectively handling multi-domain and multi-modal fake news detection through modality-specific decision contributions.
2. We significantly enhance the performance of fake news detection tasks by introducing Domain Disentanglement, a Domain-Aware Multi-View Discriminator, and a Domain-Enhanced Multi-View Decision Layer.
3. We validate the efficacy of DAMMFND through comprehensive experiments on two real-world datasets, including ablation studies to evaluate the impact of each component.
4. To the best of our knowledge, this is the first work in the field of fake news detection that incorporates domain information for multi-modal decision contributions.

Related Works

Multimodal Fake News Detection

Visual and textual information are critical in fake news detection, spurring the development of multimodal approaches. CMC (Wei et al. 2022) leverages cross-modal feature relevance to guide unimodal network training, subsequently fusing the trained features for detection. (Gao et al. 2024) augments text and visual modality decision-making through large-scale open knowledge graphs. FCINet (Bai, Liu, and Li 2024) extracts news representations from frequency, spatial, and textual domains, employing parallel cross-modal interaction to explore inter-modal dependencies. However, existing methods often simplistically fuse multimodal features, failing to effectively integrate diverse modal perspectives for comprehensive analysis.

Multi-domain Fake News Detection

The goal of multi-domain fake news detection is to achieve domain-adaptive detection by leveraging both domain information and news content. Recent approaches based on multi-domain learning have shown significant progress in this field (Fu, Peng, and Liu 2023; Chen, Fu, and Tang 2023; LI, SANG, and ZHANG 2024; Wu et al. 2024). In unimodal methods, MDFEND (Nan et al. 2021) employs a domain gating network to adapt weights of multiple text expert networks. Given that real-world news often comprises multimodal data, research has shifted towards multimodal multi-domain fake news detection. KATMF (Song et al. 2021) utilizes adversarial multi-task learning and a knowledge-enhanced Transformer to capture domain-specific feature distributions. MMDFND (Tong et al. 2024) enhances performance by modeling inter-domain commonalities and specificities through an improved PLE module. However, these methods overlook challenges such as imperfect domain labels and the differential impact of domains on various modalities.

Method

Figure 2 depicts the DAMMFND model architecture, comprising three key components: a domain information extraction and negative transfer reduction module, a multimodal feature identification module for distinguishing real and fake news, and a domain-informed decision weight generation module for different modalities.

Multi-view Features Extraction and Aggregation

Using pre-trained models, the image **I** and text **T** can be encoded into unimodal embeddings. Simultaneously, we use a set of pre-trained image-text pair encoders to encode the images and text, and then fuse this set of encoded unimodal features to obtain the multimodal features of the news. Then, we use three deep extractors to extract news representations from textual features, image features, and multimodal features respectively.

Text View Feature. Given a news text **T**, we employ a pre-trained BERT (Devlin et al. 2018) to generate its text representation e_t. Concurrently, we utilize a pre-trained CLIP (Radford et al. 2021) text encoder to obtain aligned text features $f_{\text{CLIP-T}}$. CLIP's training on a large-scale image-text dataset ensures inherent feature alignment. We then implement a Text Network (TextNET), specifically TextCNN, to extract news representations from these textual features. The feature extraction process can be expressed as:

$$r^{text} = \text{TextNET}(e_t) \tag{1}$$

Visual View Feature. For a news image **I**, we employ a pre-trained MAE (He et al. 2021) to generate its representation e_i. Simultaneously, we use a pre-trained CLIP image encoder to obtain aligned image features $f_{\text{CLIP-I}}$. An Image Network (ImageNET), implemented as a CNN in this study, extracts news representations from these visual features. The feature extraction process is expressed as:

$$r^{img} = \text{ImageNET}(e_i) \tag{2}$$

Multimodal View Feature. Multimodal representations not only encapsulate the correlation between images and text but also contain more comprehensive information. Therefore, we extract the multimodal features of the news to determine both its veracity and the domain to which the news belongs from multiple modal perspectives. Specifically, we concatenate the CLIP-encoded image and text, and then weight the concatenated features based on the correlation between the image and text to obtain the multimodal representation e_m:

$$cor = \frac{f_{\text{CLIP-T}} \cdot (f_{\text{CLIP-I}})^T}{\|f_{\text{CLIP-T}}\| \|f_{\text{CLIP-I}}\|} \tag{3}$$

$$e_m = cor \cdot \text{MLP}(f_{\text{CLIP-I}} \oplus f_{\text{CLIP-T}}) \tag{4}$$

Then, we utilize a Multimodal Network (MmNET) to extract news representations from multimodal features. In this paper, a Multilayer Perceptron (MLP) is used as MmNET, and the feature extraction process can be represented as follows:

$$r^{mm} = \text{MmNET}(e_m) \tag{5}$$

Multi-channel Features Aggregation. Inspired by multi-channel CNNs (Zheng et al. 2014; Yang et al. 2015), utilizing multi-channel TextNET, ImageNET, and MmNET to extract multiple perspective representations for each modality is beneficial. Multi-channel extractors can encourage the extractor to focus on different subspaces of representations. We can obtain multi-channel representations for the three modalities, denoted as $\{r_i^{text}\}_{i=1}^{k_{text}}$, $\{r_i^{img}\}_{i=1}^{k_{img}}$, and $\{r_i^{mm}\}_{i=1}^{k_{mm}}$, where k_{text}, k_{img}, and k_{mm} represent the number of channels for TextNET, ImageNET, and MmNET, respectively. Then, we input the text features e_t, image features e_i, and multimodal features e_m into an attention-guided gating network to perform weighted aggregation of the multiple visual perspectives of the multiple modalities. The aggregation process can be represented as follows:

$$f_{text} = \sum_{i=1}^{k_{text}} \text{softmax}(G_{text}(e_t))_i \cdot r_i^{text} \tag{6}$$

$$f_{img} = \sum_{i=1}^{k_{img}} \text{softmax}(G_{img}(e_i))_i \cdot r_i^{img} \tag{7}$$

$$f_{mm} = \sum_{i=1}^{k_{mm}} \text{softmax}(G_{mm}(e_m))_i \cdot r_i^{mm} \tag{8}$$

These features will subsequently be input into the domain disentanglement network.

Domain Disentanglement

Estimating soft masks. To effectively extract domain information, it is essential to separate domain information from news semantic information. To this end, we use an attention module that allows the aggregated features of each modality to generate two branches: one for domain information and one for news semantics. Taking the aggregated features of the text modality f_{text} as an example, we first use two MLP networks to estimate the attention scores for domain and semantic information:

$$\alpha_{dom} = \text{MLP}_{dom}(f_{text}) \tag{9}$$

$$\alpha_{se} = \text{MLP}_{se}(f_{text}) \tag{10}$$

We then use the attention scores α_{dom} and α_{se} to construct soft masks M_d and M_s, respectively. Finally, we decompose the original aggregated text features into the domain branch and the semantic branch: $f_{text}^{dom} = f_{text} \odot M_d$ and $f_{text}^{se} = f_{text} \odot M_s$.

Disentanglement. Until now, we have obtained the raw features of the domain and semantic branches. Now, we need to constrain these two branches using domain multi-label classification loss and domain uniformity loss to obtain disentangled domain features and news semantic features. The domain branch aims to capture domain features, so we perform multi-domain label classification on the representations. The supervised classification loss is defined as follows:

$$\mathcal{L}_{text}^{dom} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{K} [y_{ij} \log(\hat{y}_{ij}) + (1 - y_{ij}) \log(1 - \hat{y}_{ij})] \tag{11}$$

Where y_{ij} and $\hat{y}_{ij}$ represent the true value and predicted value of the j-th label of the i-th news, respectively. $\hat{y}_{ij}$ is obtained from the domain feature f_{text}^{dom} through a classifier. The news semantic branch aims to capture features that do

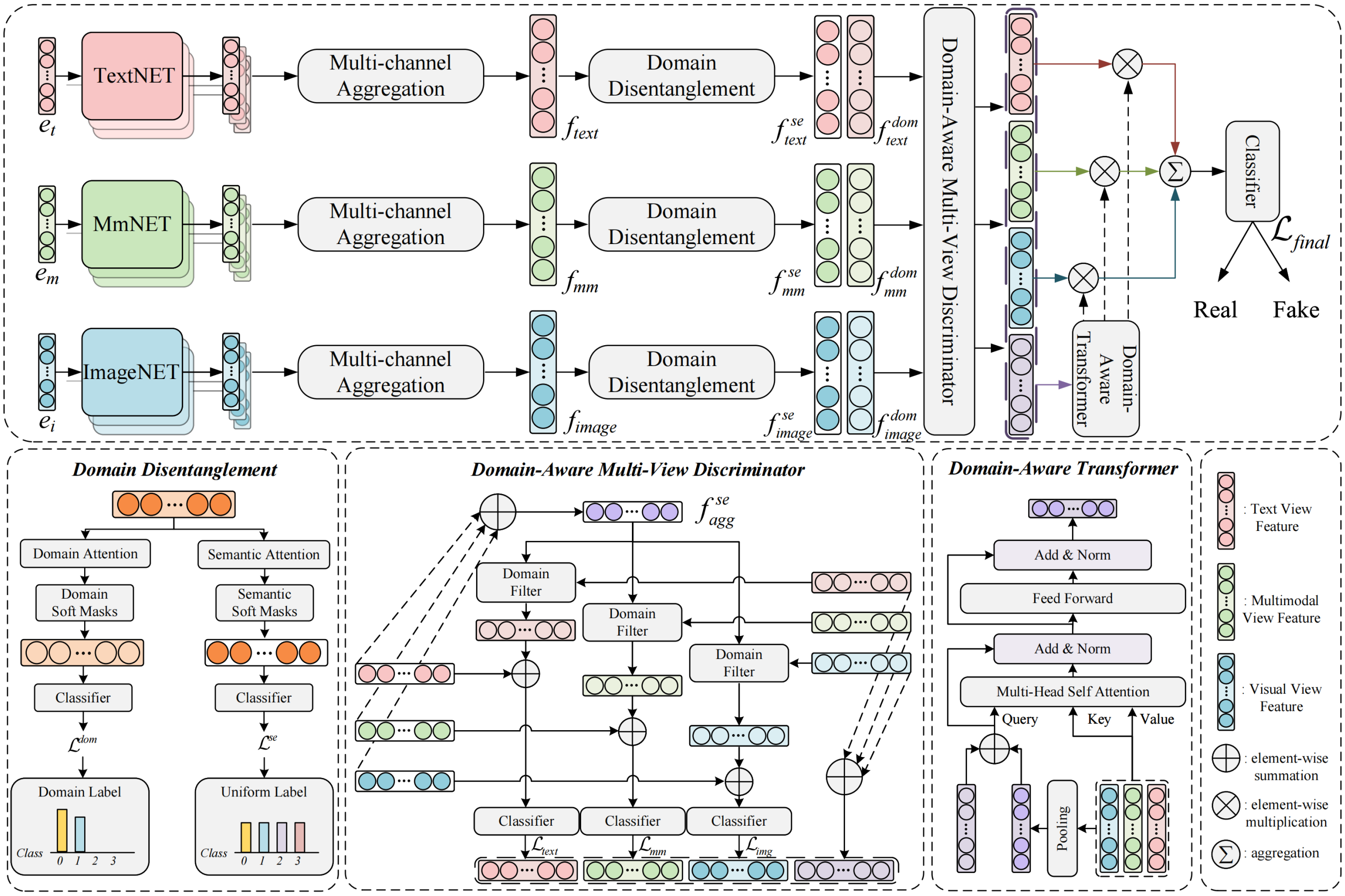


Figure 2: The network architecture of DAMMFND. TextNet, MmNet, and ImageNet are utilized to extract different modality view representations of multimodal news. Multi-channel Aggregation is employed to aggregate multi-channel features from different modality views. Domain Disentanglement extracts domain information and reduces negative transfer. The Domain-Aware Multi-View Discriminator identifies real and fake features across multimodal views. The Domain-Aware Transformer generates decision weights based on domain information. News authenticity is determined by the aggregated feature.

not contain domain information. Therefore, we encourage its predictions to be uniformly distributed across all domain categories, and define the uniform classification loss as follows:

$$\mathcal{L}^{se}_{text} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{KL}\left(y_{\mathrm{unif}}, \hat{y}_i\right) \tag{12}$$

Where KL denotes the Kullback-Leibler (KL) divergence, and y_{unif} represents the uniform distribution (i.e., equal probability for all domain categories). $\hat{y}_i$ is obtained from the semantic features f^{se}_{text} through a classifier. By constraining the domain and semantic branches with the above two loss functions, we can effectively disentangle domain features and news semantic features. Using the same approach, we can obtain the corresponding domain features and semantic features from the image aggregated features and multimodal aggregated features, which are f^{dom}_{img}, f^{se}_{img}, f^{dom}_{mm}, and f^{se}_{mm}, respectively.

Domain-Aware Multi-View Discriminator

Fake news detection typically employs three perspectives: textual, visual, and multimodal. Textual analysis examines linguistic features, sentiment, and logical consistency. Visual analysis identifies potentially manipulated or misappropriated images. Multimodal analysis evaluates semantic coherence between text and visual elements. However, news interpretations can vary significantly across domains. For example, ”virus” has distinct meanings in health, technology, and social contexts. Consequently, the effectiveness of multi-view fake news detection is inherently domain-dependent. To address this, we propose the Domain-Aware Multi-View Discriminator (DAMVD). DAMVD employs a gated network to filter domain representations relevant to fake news detection and utilizes these representations as input for each view, enabling domain-adaptive detection.

We hypothesize that not all domain features contribute equally to fake news detection. Inspired by ERASE (Jia et al. 2024), we implement gating networks for each modality representation to quantify the contribution of domain classification features to fake news detection:

$$G_i = \sigma(Gate_i(f_{text}^{se} + f_{img}^{se} + f_{mm}^{se})) \quad (13)$$

Where G_i denotes the i-th modality's gating network output, dynamically modulating the importance of modality-specific features. $Gate_i$ comprises a single fully connected layer with Sigmoid activation. Inspired by Low-Rank Adaptation (Hu et al. 2021), each gating network employs low-rank approximation to reduce dimensionality, enhancing efficiency and mitigating overfitting.

After that, we use the obtained G_i to control the retention degree of domain features $f_i^{dom} \in \{f_{text}^{dom}, f_{img}^{dom}, f_{mm}^{dom}\}$ corresponding to the i-th modality:

$$p_i^{dom} = G_i \otimes f_i^{dom} \quad (14)$$

where $p_i^{dom} \in \{p_{text}^{dom}, p_{img}^{dom}, p_{mm}^{dom}\}$ is the domain feature after filtering.

Finally, the meticulously curated domain-specific features are concatenated with the news semantic features extracted from each modality. These integrated feature vectors are subsequently fed into their respective modality-specific prediction layers to obtain prediction scores from each modal perspective: $\hat{y}_{text}$ for textual content, $\hat{y}_{img}$ for visual elements, and $\hat{y}_{mm}$ for multimodal integration. This approach leverages the complementary strengths of different modalities, enhancing the model's ability to detect nuanced patterns indicative of fake news across various information channels.

Domain-Enhanced Multi-view Decision Layer

Fake news detection exhibits significant modality-specific biases across domains, impacting multi-view decision-making efficacy. For instance, images may carry more weight in entertainment news, while textual accuracy is crucial for political and economic reports. In social domains, even factual content can be misleading when paired with controversial imagery. Thus, investigating domain influence on multi-view decision-making is critical. To quantify view contributions in domain-specific fake news detection, we propose a Domain-Enhanced Multi-view Decision Layer. This module's core is the Domain-aware Transformer (DAFormer), which integrates domain information for multi-view weight prediction.

First, we apply average pooling to the representations of three modalities (text f_{text}^{se}, image f_{img}^{se}, fusion f_{mm}^{se}) to obtain global view representations f_{agg}^{se}, enabling each view to perceive the global information of the current decision:

$$f_{agg}^{se} = \text{Mean}(f_{text}^{se}, f_{img}^{se}, f_{mm}^{se}) \quad (15)$$

Next, we add f_{agg}^{se} and the domain features $\{p_{text}^{dom}, p_{img}^{dom}, p_{mm}^{dom}\}$ to form the Q, ugmenting the decision weight generation with domain-specific information:

$$Q = \text{Add}(f_{agg}^{se}, p_{text}^{dom}, p_{img}^{dom}, p_{mm}^{dom}) \quad (16)$$

We then employ a Transformer(Vaswani et al. 2017) to process the multi-view information. In details, we use Q as the query and $[f_{text}^{se}, f_{img}^{se}, f_{mm}^{se}]$ as K and V for the attention mechanism, performing the following calculation:

$$\begin{aligned} X_1 &= \text{LayerNorm}(Q + \text{MultiHeadAttention}(Q, K, V)) \\ X_2 &= \text{LayerNorm}(X_1 + \text{FFN}(X_1)) \end{aligned} \quad (17)$$

where the output X_2 from the Transformer is then passed through a dimension-3 output layer followed by a softmax function to obtain the weights for each view:

$$[w_{text}, w_{img}, w_{mm}] = \text{softmax}(W_o X_2 + b_o) \quad (18)$$

where $W_o \in \mathbb{R}^{3\times d}$ and $b_o \in \mathbb{R}^3$ are learnable parameters.

Finally, we obtain the ultimate fake news prediction $\hat{y}$ by applying weighted aggregation to the predictions from the three views using their respective weights:

$$\hat{y} = (w_{text} \cdot \hat{y}_{text} + w_{img} \cdot \hat{y}_{img} + w_{mm} \cdot \hat{y}_{mm})/3 \quad (19)$$

Loss Function

To enhance the discriminative power of each view in fake news detection, we compute the BCE loss between the predicted probabilities and ground truth labels for textual ($\mathcal{L}_{text}$), visual ($\mathcal{L}_{img}$), and multimodal ($\mathcal{L}_{mm}$) views. We also calculate the BCE loss ($\mathcal{L}_{final}$) between the final prediction $\hat{y}$ and the true label y. Additionally, we define domain disentanglement losses $\mathcal{L}_t^{dd}$, $\mathcal{L}_i^{dd}$, and $\mathcal{L}_{mm}^{dd}$ for textual, visual, and multimodal features, respectively. The total loss for DAMMFND is thus:

$$\begin{aligned} \mathcal{L} = \mathcal{L}_{final} + \alpha_{fnd}(\mathcal{L}_{text} + \mathcal{L}_{img} + \mathcal{L}_{mm}) + \alpha_{dom}(\mathcal{L}_{text}^{dom} \\ + \mathcal{L}_{img}^{dom} + \mathcal{L}_{mm}^{dom}) + \alpha_{se}(\mathcal{L}_{text}^{se} + \mathcal{L}_{img}^{se} + \mathcal{L}_{mm}^{se}) \end{aligned} \quad (20)$$

where α_{fnd} is the weight of multi-view tasks, α_{dom} is the weight of domain-supervised classification loss, and α_{se} is the weight of domain-uniform classification loss.

Experiments

Experimental Settings

Datasets We evaluated our model using two real-world datasets: Weibo (Wang et al. 2018) and Weibo-21 (Nan et al. 2021). The Weibo dataset contains 7,532 news articles (3,749 true, 3,783 fake) for training and 1,996 articles (996 true, 1,000 fake) for testing. We processed this dataset according to benchmark standards and categorized it into nine domains: finance, health, military, science, politics, international, education, entertainment, and society to facilitate domain adaptation. Weibo21, a multi-domain dataset, comprises 9,127 articles (4,640 true, 4,487 fake), which we partitioned into training and test sets following established benchmark procedures. Table 1 presents the statistical breakdown of both datasets.

Implementation Details. In the text data encoding section, we set 197 as the maximum length for text input and utilized pre-trained BERT (Devlin et al. 2018) and CLIP models for text encoding. For the visual data encoding, we first resized the input images to 224 × 224 pixels and employed pre-trained MAE (He et al. 2021) and CLIP models to encode the image data. Accuracy, precision, and F1 score are adopted as the evaluation metrics for our model. In the DAMMFND framework's loss formula (Eq. 20), the parameter α is set to 0.25. We set the number of channels k_{text}, k_{img} and k_{mm} to 18. All codes are developed using PyTorch (Paszke et al. 2019) and executed on an NVIDIA RTX 4090 graphics processing unit.

	Method	Sci.	Mil.	Edu.	Soc.	Pol.	Hlth.	Fin.	Ent.	Dis./Int	Overall		
											F1	Acc	AUC
Weibo	MMoE*	0.578	0.911	0.851	0.885	0.735	0.826	0.813	0.830	0.883	0.874	0.874	0.950
	MoSE*	0.793	0.738	0.834	0.912	0.764	0.859	0.791	0.844	0.883	0.890	0.890	0.954
	MDFEND*	0.774	0.911	0.897	0.902	0.763	0.878	0.808	0.881	0.874	0.904	0.904	0.965
	M^{3}DFEND*	0.792	0.903	0.923	0.912	0.765	0.863	0.899	0.899	0.876	0.928	0.928	0.969
	KATMF	0.831	0.908	0.924	0.895	**0.823**	0.898	0.903	0.904	0.894	0.929	0.930	0.969
	MMDFND*	0.824	0.911	0.941	0.939	0.735	0.913	0.917	0.917	0.888	0.934	0.934	0.972
	DAMMFND	**0.853**	**0.911**	**0.956**	**0.943**	0.822	**0.939**	**0.937**	**0.956**	**0.928**	**0.943**	**0.944**	**0.983**
Weibo-21	MMoE*	0.875	0.911	0.870	0.875	0.862	0.936	0.856	0.888	0.877	0.894	0.894	0.954
	MoSE*	0.850	0.885	0.881	0.872	0.880	0.917	0.867	0.891	0.867	0.893	0.894	0.954
	MDFEND*	0.830	0.938	0.891	0.898	0.886	0.940	0.895	0.906	0.900	0.913	0.913	0.970
	M^{3}DFEND*	0.829	0.950	0.899	0.908	0.882	**0.946**	0.900	0.931	0.889	0.921	0.921	0.975
	KATMF	0.914	0.928	0.913	0.895	0.902	0.914	0.871	0.937	0.898	0.923	0.928	0.975
	MMDFND*	**0.937**	0.953	0.852	0.945	0.965	0.920	0.884	0.959	0.919	0.939	0.939	0.977
	DAMMFND	0.932	**0.953**	**0.932**	**0.917**	**0.982**	0.919	**0.941**	**0.990**	**0.945**	**0.947**	**0.947**	**0.985**

Table 2: Comparison between DAMMFND and the latest multi-domain fake news detection methods on Weibo and Weibo-21. *: open-source.

	Method	Accuracy	F1 score	
			Real News	Fake News
Weibo	SpotFake*	0.892	0.932	0.739
	CAFE*	0.840	0.842	0.837
	CMC	0.893	0.899	0.907
	BMR*	0.918	0.914	0.904
	MRHFR*	0.922	0.918	0.910
	DAMMFND	**0.944**	**0.945**	**0.943**
Weibo-21	SpotFake*	0.851	0.828	0.866
	CAFE*	0.882	0.885	0.876
	CMC*	0.897	0.903	0.912
	BMR*	0.929	0.927	0.925
	MRHFR*	0.932	0.928	0.926
	DAMMFND	**0.947**	**0.948**	**0.947**

Table 3: Comparison between DAMMFND and the latest multimodal fake news detection methods on Weibo and Weibo-21. *: open-source.

Baselines. We categorize our baselines into three groups. The first group is single modal multi-domain methods, including: *MMoE* (Ma et al. 2018), *MoSE* (Qin et al. 2020), *MDFEND* (Nan et al. 2021) and *M^{3}DFEND* (Zhu et al. 2022b). The second group is multimodal multi-domain methods, including: *KATMF* (Song et al. 2021) and *MMDFND* (Tong et al. 2024). The third group is multimodal single domain methods, including: *SpotFake* (Singhal et al. 2019), *CAFE* (Chen et al. 2022), *CMC* (Wei et al. 2022), *BMR* (Ying et al. 2023) and *MRHFR* (Wu, Liu, and Zhang 2023).

Overall Performance

To rigorously evaluate the effectiveness of our proposed DAMMFND framework, we conducted extensive experiments using the Weibo and Weibo21 datasets, and compared the performance of our model with the aforementioned three categories of baseline models. We present the comparative results with the first two categories in Table 2 and the third category in Table 3. As demonstrated in Table 2, DAMMFND significantly outperforms most multi-domain models in terms of F1-scores across multiple domains and overall. Table 3 illustrates that DAMMFND exhibits substantial improvements in accuracy and F1-scores compared to most multimodal methods. Specifically, DAMMFND achieves state-of-the-art results with an accuracy of **94.4%** on the Weibo dataset and **94.7%** on the Weibo21 dataset. From these results, we can draw the following conclusions:

- Among multi-domain methods in Table 2, KATMF and MMDFND outperform MDFEND and M^{3}DFEND, demonstrating the superiority of multimodal multi-domain approaches in capturing cross-modal semantic discrepancies. MDFEND and M^{3}DFEND generally surpass MMoE and MoSE across domains. While MMoE and MoSE employ shared experts to leverage cross-domain knowledge without domain labels, their rigid sharing fails to extract domain-specific information effectively. In contrast, MDFEND and M^{3}DFEND utilize domain labels and implement domain-adaptive gating networks via domain embeddings, yielding superior performance among baselines.
- As shown in Table 3, in multimodal methods, SpotFake excels in rumor classification by utilizing pre-trained models for unimodal feature extraction. CAFE enhances multimodal fusion through semantic space alignment, while CMC and BMR effectively leverage cross-modal correlations. MRHFR outperforms other baselines by modeling users' reading habits and exploring semantic discrepancies between unimodal and multimodal features. However, these methods fail to capture cross-domain knowledge and domain-specific insights, resulting in suboptimal performance on cross-modal datasets compared to multi-domain multimodal approaches.

	Method	Accuracy	F1 score	
			Real	Fake
Weibo	**DAMMFND**	**0.944**	**0.945**	**0.943**
	-w/o MCFA	0.934	0.936	0.928
	-w/o DD	0.927	0.928	0.926
	-w/o DAMVD(gate)	0.931	0.933	0.930
	-w/o DAMVD(views)	0.928	0.932	0.929
	-w/o DEMD	0.924	0.927	0.922
Weibo-21	**DAMMFND**	**0.947**	**0.948**	**0.947**
	-w/o MCFA	0.938	0.940	0.937
	-w/o DD	0.931	0.933	0.931
	-w/o DAMVD(gate)	0.934	0.937	0.932
	-w/o DAMVD(views)	0.932	0.934	0.928
	-w/o DEMD	0.922	0.920	0.923

Table 4: Ablation study on the network design of DAMMFND on two datasets.

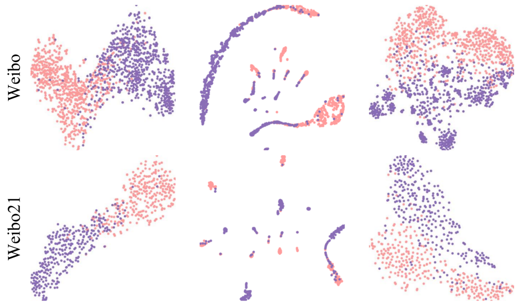
- DAMMFND's superior performance can be attributed to: (1) enhanced multimodal representations through Multi-view Features Extraction and Aggregation; (2) improved domain information capture and reduced negative transfer while facilitating cross-domain information sharing; and (3) effective characterization of domain-aware multi-view decision-making, modeling the logical process of fake news discrimination across domains.
- Figure 3 presents t-SNE visualizations of features learned by DAMMFND, MMDFND, and MDFEND on Weibo and Weibo21 test sets. DAMMFND generates fewer outliers in fake news representations and demonstrates reduced overlap between real and fake news embeddings compared to MMDFND and MDFEND. These findings further substantiate DAMMFND's superior performance in multimodal fake news detection.

Ablation Study

To demonstrate the effectiveness of each component in the DEFND framework, we conducted an ablation study by systematically removing each component and performing comparative analysis. Specifically, we employed the following experimental settings: (1) **w/o MCFA**: removing the Multi-channel Features Aggregation, directly using the original text features, visual features, and multimodal features. (2) **w/o DD**: removing the Domain Disentanglement module, eliminating the two domain disentanglement-related losses. (3) **w/o DAMVD(gate)**: removing the domain information filtering mechanism in the Domain-Aware Multi-View Discriminator. (4) **w/o DAMVD(views)**: removing the three modality views in the Domain-Aware Multi-View Discriminator, using only the final overall prediction result. (5) **w/o DEMD**: removing the Domain-Enhanced Multi-view Decision Layer, simply averaging the predictions from different views to achieve multi-view decision-making.

Table 4 presents our ablation study results. The original DAMMFND outperforms all variants, demonstrating each component's effectiveness. Key findings include:

- Removing Multi-channel Features Aggregation significantly decreases performance, highlighting the importance of extracting and aggregating multimodal features for enhanced representations.
- Eliminating domain disentanglement substantially reduces performance, confirming its role in accurately extracting domain-specific information and mitigating negative transfer.
- Removing the domain information filtering mechanism from the Domain-Aware Multi-View Discriminator led to F1-score(for genuine news, same below) decreases of 1.2% and 1.1% on the two datasets, respectively. This indicates that not all domain features contribute positively to fake news detection, and their indiscriminate inclusion may lead to overfitting.
- Omitting the three modal views in the Domain-Aware Multi-View Discriminator causes F1 score decreases of 1.3% and 1.4%. This validates that multi-view discrimination enhances the model's capacity to detect subtle patterns across various information channels for fake news identification.
- Removing the Domain-Enhanced Multi-view Decision Layer decreases accuracy by 2.5% and 2.8% on the two datasets. This highlights the domain-specific reliance on different modalities and validates our approach's ability to quantify modal contributions, thereby improving overall fake news detection performance.



(a) DAMMFND (b) MMDFND (c) MDFEND

Figure 3: T-SNE visualization of test set features. Same color dots indicate the same label.

Conclusion

This paper proposes the Domain-Aware Multi-Modal Multi-View Fake News Detection (DAMMFND) framework. DAMMFND effectively extracts more accurate domain information through Domain Disentanglement, while simultaneously mitigating negative transfer between domains. Additionally, DAMMFND introduces a Domain-Aware Multi-View Discriminator and a Domain-Enhanced Multi-view Decision Layer, which accurately quantify the contribution of domain information to multimodal, multi-view decision-making processes. Extensive experiments conducted on two real-world benchmark datasets demonstrate that the proposed model outperforms state-of-the-art baselines.

References

Bai, Y.; Liu, Y.; and Li, Y. 2024. Learning Frequency-Aware Cross-Modal Interaction for Multimodal Fake News Detection. *IEEE Transactions on Computational Social Systems*.

Chen, Y.; Li, D.; Zhang, P.; Sui, J.; Lv, Q.; Tun, L.; and Shang, L. 2022. Cross-modal ambiguity learning for multimodal fake news detection. In *Proceedings of the ACM web conference 2022*, 2897–2905.

Chen, Z.; Fu, C.; and Tang, X. 2023. Multi-domain Fake News Detection with Fuzzy Labels. In *International Conference on Database Systems for Advanced Applications*, 331–343. Springer.

Cui, L.; Shu, K.; Wang, S.; Lee, D.; and Liu, H. 2019. defend: A system for explainable fake news detection. In *Proceedings of the 28th ACM international conference on information and knowledge management*, 2961–2964.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Fu, L.; Peng, H.; and Liu, S. 2023. KG-MFEND: an efficient knowledge graph-based model for multi-domain fake news detection. *The Journal of Supercomputing*, 79(16): 18417–18444.

Gao, X.; Wang, X.; Chen, Z.; Zhou, W.; and Hoi, S. C. 2024. Knowledge enhanced vision and language model for multimodal fake news detection. *IEEE Transactions on Multimedia*.

He, K.; Chen, X.; Xie, S.; Li, Y.; Dollár, P.; and Girshick, R. 2021. Masked Autoencoders Are Scalable Vision Learners. *arXiv:2111.06377*.

Hu, E. J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; and Chen, W. 2021. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*.

Jia, P.; Wang, Y.; Du, Z.; Zhao, X.; Wang, Y.; Chen, B.; Wang, W.; Guo, H.; and Tang, R. 2024. ERASE: Benchmarking Feature Selection Methods for Deep Recommender Systems. *arXiv preprint arXiv:2403.12660*.

Jing, J.; Wu, H.; Sun, J.; Fang, X.; and Zhang, H. 2023. Multimodal fake news detection via progressive fusion networks. *Information processing & management*, 60(1): 103120.

Li, H.; Yang, Z.; Ma, Y.; Bin, Y.; Yang, Y.; and Chua, T.-S. 2024a. MM-Forecast: A Multimodal Approach to Temporal Event Forecasting with Large Language Models. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 2776–2785.

Li, J.; Bin, Y.; Ma, Y.; Yang, Y.; Huang, Z.; and Chua, T.-S. 2024b. Filter-based Stance Network for Rumor Verification. *ACM Transactions on Information Systems*, 42(4): 1–28.

Li, J.; Bin, Y.; Peng, L.; Yang, Y.; Li, Y.; Jin, H.; and Huang, Z. 2024c. Focusing on Relevant Responses for Multi-modal Rumor Detection. *IEEE Transactions on Knowledge and Data Engineering*.

LI, J.; SANG, G.; and ZHANG, Y. 2024. APK-CNN and Transformer-enhanced multi-domain fake news detection model. *Journal of Computer Applications*, 0.

Ma, J.; Zhao, Z.; Yi, X.; Chen, J.; Hong, L.; and Chi, E. H. 2018. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, 1930–1939.

Nan, Q.; Cao, J.; Zhu, Y.; Wang, Y.; and Li, J. 2021. MDFEND: Multi-domain fake news detection. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 3343–3347.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.

Qi, P.; Cao, J.; Yang, T.; Guo, J.; and Li, J. 2019. Exploiting multi-domain visual information for fake news detection. In *2019 IEEE international conference on data mining (ICDM)*, 518–527. IEEE.

Qin, Z.; Cheng, Y.; Zhao, Z.; Chen, Z.; Metzler, D.; and Qin, J. 2020. Multitask mixture of sequential experts for user activity streams. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 3083–3091.

Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, 8748–8763. PMLR.

Singhal, S.; Shah, R. R.; Chakraborty, T.; Kumaraguru, P.; and Satoh, S. 2019. Spotfake: A multi-modal framework for fake news detection. In *2019 IEEE fifth international conference on multimedia big data (BigMM)*, 39–47. IEEE.

Song, C.; Ning, N.; Zhang, Y.; and Wu, B. 2021. Knowledge augmented transformer for adversarial multidomain multiclassification multimodal fake news detection. *Neurocomputing*, 462: 88–100.

Tong, Y.; Lu, W.; Zhao, Z.; Lai, S.; and Shi, T. 2024. MMDFND: Multi-modal Multi-Domain Fake News Detection. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 1178–1186.

Van der Maaten, L.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of machine learning research*, 9(11).

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017. Attention Is All You Need. arXiv:1706.03762.

Wang, Y.; Ma, F.; Jin, Z.; Yuan, Y.; Xun, G.; Jha, K.; Su, L.; and Gao, J. 2018. Eann: Event adversarial neural networks for multi-modal fake news detection. In *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining*, 849–857.

Wei, Z.; Pan, H.; Qiao, L.; Niu, X.; Dong, P.; and Li, D. 2022. Cross-modal knowledge distillation in multi-modal fake news detection. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4733–4737. IEEE.

Wu, D.; Tan, Z.; Zhao, H.; Jiang, T.; and Qi, M. 2024. LIMFA: label-irrelevant multi-domain feature alignment-based fake news detection for unseen domain. *Neural Computing and Applications*, 36(10): 5197–5215.

Wu, L.; Liu, P.; and Zhang, Y. 2023. See how you read? multi-reading habits fusion reasoning for multi-modal fake news detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 13736–13744.

Yang, J.; Nguyen, M. N.; San, P. P.; Li, X.; and Krishnaswamy, S. 2015. Deep convolutional neural networks on multichannel time series for human activity recognition. In *Ijcai*, volume 15, 3995–4001. Buenos Aires, Argentina.

Ying, Q.; Hu, X.; Zhou, Y.; Qian, Z.; Zeng, D.; and Ge, S. 2023. Bootstrapping multi-view representations for fake news detection. In *Proceedings of the AAAI conference on Artificial Intelligence*, volume 37, 5384–5392.

Zhang, W.; Deng, L.; Zhang, L.; and Wu, D. 2022. A survey on negative transfer. *IEEE/CAA Journal of Automatica Sinica*, 10(2): 305–329.

Zhang, Y.; and Wallace, B. 2015. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. *arXiv preprint arXiv:1510.03820*.

Zheng, Y.; Liu, Q.; Chen, E.; Ge, Y.; and Zhao, J. L. 2014. Time series classification using multi-channels deep convolutional neural networks. In *International conference on web-age information management*, 298–310. Springer.

Zhou, X.; Cao, J.; Jin, Z.; Xie, F.; Su, Y.; Chu, D.; Cao, X.; and Zhang, J. 2015. Real-time news cer tification system on sina weibo. In *Proceedings of the 24th international conference on world wide web*, 983–988.

Zhu, Y.; Sheng, Q.; Cao, J.; Li, S.; Wang, D.; and Zhuang, F. 2022a. Generalizing to the future: Mitigating entity bias in fake news detection. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2120–2125.

Zhu, Y.; Sheng, Q.; Cao, J.; Nan, Q.; Shu, K.; Wu, M.; Wang, J.; and Zhuang, F. 2022b. Memory-guided multi-view multi-domain fake news detection. *IEEE Transactions on Knowledge and Data Engineering*.