J. Integr. Sci. Technol. 2025, 13(6), 1138

Journal of Integrated
SCIENCE & TECHNOLOGY

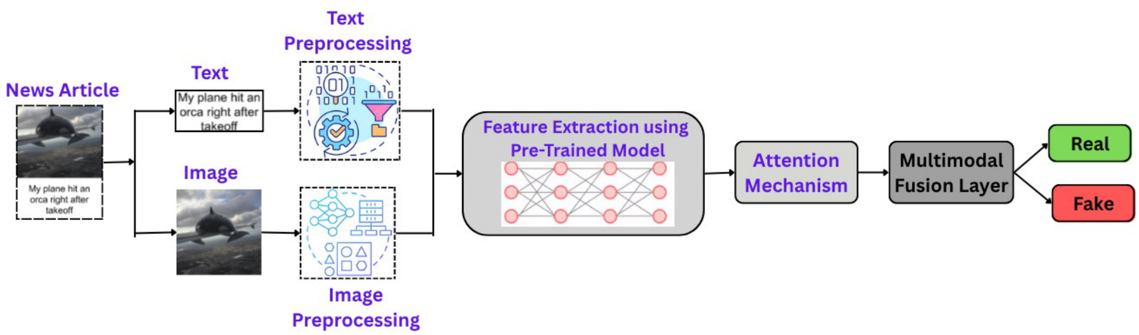
MTLFND: Multimodal fake news detection using attention mechanism and transfer learning

Sudha Patel[1], Shivangi Surati[2]

[1]Kadi Sarva Vishwavidyalaya, Gandhinagar, Gujarat, India. [2]Department of Computer Science and Engineering, School of Technology, Pandit Deendayal Energy University, Gandhinagar, Gujarat, India

Submitted on: 03-Sep-2024, Accepted and Published on: 15-May-2025 Article

ABSTRACT



In the era of increasing information, fighting fake data has become a topmost significance. The Multimodal Transfer Learning framework for False News Detection (MTLFND) is suggested in the paper, presenting an innovative approach to enhance accuracy and robustness in recognizing misleading information across diverse media formats. While existing approaches rely on performing multimodal fusion techniques, the proposed method addresses this by introducing a novel MTLFND module. The cross-modal similarity between textual and graphical features is captured by this module, intelligently leveraging pre-trained knowledge. The extracted features are dynamically weighed and combined with this similarity information, with relevant details from both modalities being prioritized during classification. This is the first state-of-the-art method to employ redundancy reduction and modality-wise attention to further refine the multimodal features before feeding them into the final classifier according to the best of our knowledge. Numerous experiments show that the suggested model is successful, surpassing other leading methods in accuracy across various datasets. The value of multimodal transfer learning models for enhanced flexibility in feature selection is highlighted by this advancement, paving a way for further research in the promising path.

Keywords: Neural Networks, Fake News Detection, Multimodal Fake News Detection, Attention Mechanism, Multimodality

INTRODUCTION

Today's digital era emphasizes the importance of detecting false news. Online social networks have surpassed traditional mediums like newspapers and magazines, enabling users to connect with others and share opinions.[1,2] But this change has also hastened the spread of false information, which can easily be manipulated and circulated, resulting in misinformation and manipulation of public opinion.[3–5] Detecting fake news manually is impractical. Hence, the importance of automated approaches using machine learning or deep learning plays a crucial role.[6] Fake news are normaly an altered form of credible news that is intended tio mislead , create uncertainty and spread fastly.[7,8] The approaches based on the content can be classified into their inherent content mainly on news text.[9] Initial attempts to detect fake news primarily concentrated on analyzing text or image content in isolation.[10,11] These methods often relied upon pre-trained models to assess the logical and semantic coherence of input data while considering factors like grammatical errors or image manipulation traces[12] There has been significant development with multimodal fake news detection techniques based on semantic details.[13,14] However, as news articles and posts increasingly contain huge data across multiple modalities, like images, text, comments, and social engagement metrics, there's a need to explore the correlation between these modalities for more effective detection.[15,16] Cross Model knowledge is crucial to enable fake news detection precisely.[17,18]

Most in recent times, there has been a substantial rise in research endeavours focused on combining multimodal features to identify

*Corresponding Author: Shivangi Surati, Department of Computer Science and Engineering, School of Technology, Pandit Deendayal Energy University, Gandhinagar, Gujarat, India. Tel: +917923275360
Email: shivangi.surati@gmail.com

Cite as: *J. Integr. Sci. Technol., 2025,13(6), 1138.*
DOI: 10.62110/sciencein.jist.2025.v13.1138

irregular messages on social media and in the news.[19,20] Apart from integrating features extracted from images and texts, researchers often incorporate additional modalities such as utilising spreading graphs, support percentages, and comments to evaluate the authenticity of postings.[21] While these additional modalities provide interactive insights, they pose challenges compared to static modalities like images and texts.[22,23] Interactive modalities are less reliable due to their dynamic nature.[24,25] For instance, past actions in the news articles that recently registered users have submitted may not provide any clues, and comments or up-votes may change over time, potentially leading to varied forensic results. Consequently, there's a growing interest in revisiting Fake News Detection (FND) approaches that solely rely on static modalities.[26,27]

Despite recent progress, there remain unresolved challenges in multimodal FND that demand attention:

- Determining effective calculation of feature similarity across different modalities and its impact on FND decision-making.[28]
- Uncertainty regarding the efficacy of methods like Variational AutoEncoders (VAEs) in minimizing divergence of Kullback-Leibler (KL) for paired image-text combination.[12]
- Lack of clarity on the necessity of models like Multimodal Variational AutoEncoder (MVAE) in the FND context [12].
- Underexplored utilization of more sophisticated multimodal learning paradigms and pre-trained models.[29]
- Despite the popularity of multiple modalities, many algorithms focus on static modalities due to their stability and reliability.[30,31]

Hence, to overcome above limitations, MTLFND- a multimodal fake news detection model that utilizes pre-trained models along with an attention mechanism, is introduced in this paper. Multitask learning explicitly assesses the relationship between text and images in targeted posts, guiding feature fusion, and decision-making stages.[32–34] Images are encoded using fine-tuned ResNet and pre-trained image modules, while text is encoded using fine-tuned ROBERTA and text modules. Unimodal features are generated by combining these features with their fine-tuned counterparts. Combined features comprising the two outputs undergo size reduction through projection heads to distill relevant features for fake news detection. Additionally, cosine similarity measures the relationship between text and image, regulating the contribution of combined features.[35,36] The relevance of these characteristics in identifying bogus news is adaptively measured by scores produced by an attention layer, followed by a classifier that distinguishes bogus news from legitimate sources based on the summarized features.[37,38]

Building upon the previous proposed work MTL-rtFND,[39] this research presents an enhanced multimodal fake news detection model with novel features. The Paper summarized the contributions follows:

- Incorporation of textual and visual features with an improved performance.
- Utilization of multimodal transfer learning and attention mechanism to address cross-modal ambiguity.
- Utilization of a fine-tuneable ResNet module alongside a pre-trained image module for image encoding.
- Encoding of textual content through a fine-tunable ROBERTA and separate text modules.
- Fusion of unimodal features to create combined features through multimodal transfer learning for fake detection outputs.

MTLFND employs three heads for projection to handle mixed and unimodal characteristics separately, effectively reducing the dimensions and distilling the essential elements for identifying false news. The suggested architecture computes the attention score of MTLFND outputs. Then it standardizes it to derive the cross-modal similarity score, that assists in adjusting the combined feature weights based on the correlation between image and text. Moreover, within MTLFND, an attention layer is incorporated, which generates three scores to dynamically evaluate the importance of characteristics of their role towards bogus news detection. Ultimately, a classifier uses streamlined features to differentiate between bogus and genuine news, present a holistic method to detecting multimodal fake information.

DESIGNED METHOD

Overview of the approach

This work presents advanced multimodal feature learning models as a novel approach to address the challenges and limitations of present methods in multimodal fake news recognition. The main limitations of existing methods are that they rely on the mechanisms of tailored attention for guiding the fusion of unimodal features. However, it raises questions regarding the calculation of feature similarity and its impact on the decision-making procedure in fake news recognition. Furthermore, while pre-trained multimodal feature learning models hold promise, their potential has not been fully exploited in existing approaches.

However, our proposed method, MTLFND, introduces a comprehensive framework for identifying bogus news that effectively addresses the limitations of existing methods. The model utilizes transfer learning and multimodal feature extraction to enhance its proficiency in mining vital features for fake news recognition. The following are the main features of the suggested MTLFND approach:

- **Integration of Transfer Learning:** The proposed method harnesses transfer learning to extract rich representations using text and image modalities, utilizing the encoder based on ResNet for images, an encoder based on ROBERTA for text, and two CLIP encoders in pairs.
- **Efficient Fusion Strategy:** MTLFND incorporates a novel fusion strategy by concatenating features produced by CLIP that are weighted according to standardized cross-modal similarity, efficiently capturing supplementary data from both modalities.
- **Modality-wise Attention Module:** It introduces the attention module according to a modality that reweights adaptively and aggregates features, improving the model's capability to discern crucial data for fake news recognition.

- **Redundancy Reduction:** Features extracted from both modalities are further processed to reduce redundancy before being input into the ultimate classification, improving the efficacy and effectiveness of the detection process.

Proposed Model Architecture

The MTLFND network architecture is presented in Figure 1. There are four basic components in the pipeline: the single-mode characteristic module, the MTLFND-based module, the attention and projection module, and the classifier.

The architecture retrieves the feature by using an already-trained ROBERTA model. $f_{\mathrm{ROBERTA}} \in \mathbb{R}^{n_{BERT}}$ from the text input x_{text} resulting in a vector $f_{\mathrm{ROBERTA}} \in \mathbb{R}^{n_{BERT}}$. Similarly, for image input x_{image}, it employs ResNet to generate deep representation such as sub-ResNet, $f_{\mathrm{ResNet}} \in \mathbb{R}^{n_{\mathrm{ResNet}}}$. Additionally, it incorporates MTLFND modules to encode image and text inputs, yielding features $f_{\mathrm{MTLFND-T}} \in \mathbb{R}^{n_{\mathrm{MTLFND}}}$ and $f_{\mathrm{MTLFND-I}} \in \mathbb{R}^{n_{\mathrm{MTLFND}}}$, respectively.

Classification and goal function: To forecast the label $\hat{y}$, the aggregated representation mAgg is fed into a two-layer fully connected network as the classifier Fcls The goal of MTLFND is to accurately forecast which news is fraudulent by minimizing the cross-entropy loss:

$$L_{CE} = y\log(\hat{y}) + (1-y)\log(1-\hat{y})$$

Mathematical model

The suggested multimodal fake news recognition using a transfer learning framework incorporates a comprehensive mathematical model to effectively combine features extracted from image and text modalities. The mathematical formulation for fake news detection is presented in detail. It outlines the feature extraction process, the incorporation of the attention mechanism, the application of CNN processing, and the use of a SoftMax layer in this paper. The mathematical model serves as a foundational framework for implementing MTLFND, providing insights into the underlying mechanisms driving the fusion of multimodal features and the decision-making process in identifying fake news. By elucidating the mathematical underpinnings of the proposed method, it aims to offer a clear and structured understanding of MTLFND's functionality and efficacy in addressing the multimodal fake news detection challenges.

Step 1: Image and Text Feature Extraction:

Let I denote the image features, and T denotes the text features.

- Q_T, K_T, and V_T as query vectors, key vectors, and value vectors for text features.
- Q_I, K_I, and V_I are query vectors, key vectors, and value vectors for image features.
- α_T and α_I as attention scores for textual and visual features.
- C_T and C_I are context vectors for textual and visual features.

Step 2: Attention Mechanism:

1. Query, Key, and Value Vectors:
   - Q_T and Q_I are the text and image characteristics query vectors, respectively.
   - Identify the key vectors for image features (K_I) and text features (K_T).
   - Identify V_T and V_I, the value vectors for text and image characteristics, respectively.
2. Attention Scores:

For text features, compute attention scores αT, and for image features, compute αI:

$$\alpha T = \mathrm{softmax}(Q_T K_T^T) \quad (1)$$

$$\alpha I = \mathrm{softmax}\ (Q_I K_T^T) \quad (2)$$

3. Weighted Sum:

Compute context vectors C_T and C_I by taking the value vectors' weighted sum using attention scores:

$$C_T = \sum_{j=1}^{N} \alpha_{Tj} V_{Tj} \quad (3)$$

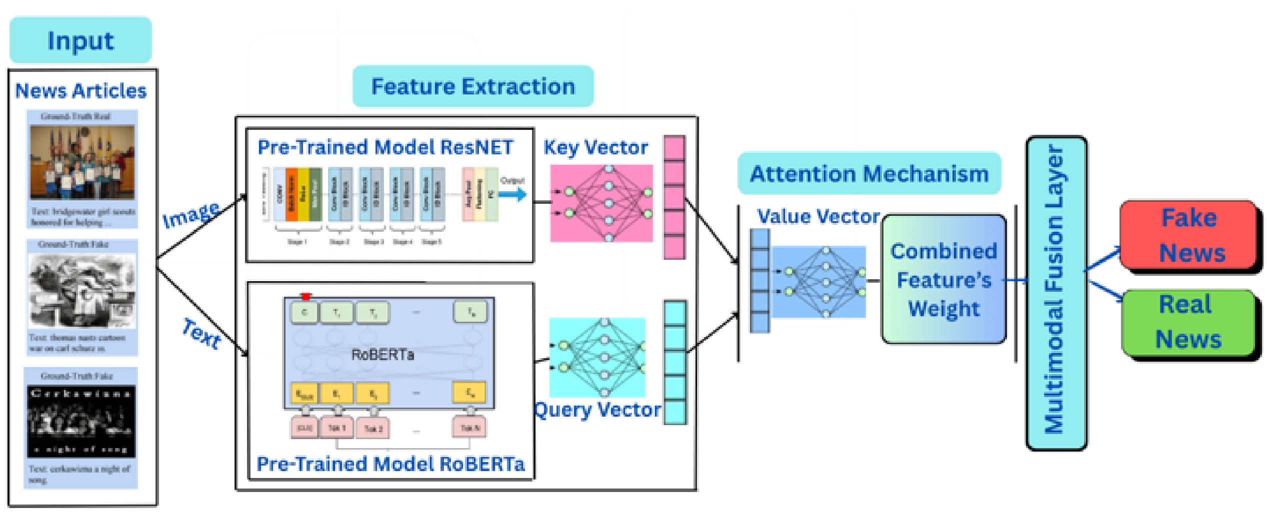


Figure 1. The proposed MTLFND architecture.

$$C_I = \sum_{j=1}^{N} \alpha_{Ij} V_{Ij} \quad (4)$$

Step 3: CNN Processing:

Concatenate or merge context vectors C_T and C_I with original image features I and text features T.

$$\text{Input to CNN} = \text{Concatenate}(C_T, C_I, I, T) \quad (5)$$

Step 4: SoftMax Layer:

Input the concatenated features into a fully connected layer followed by SoftMax activation to get predicted probabilities for fake news recognition.

Thus, Section 2 emphasizes a mathematical model of the proposed model, and Section 3 will discuss its implementation.

EXPERIMENTS

Datasets

The social media datasets are used in this research for the experiments Weibo, Gossipcop, and PolitiFact. It preprocesses the data by filtering out unimodal news posts lacking text or image descriptions. One image is chosen randomly from a group of related images in a news post. In particular, the training set of the widely used Chinese dataset Weibo comprises 1,996 news items, whereas the test set comprises 3,749 true news and 3,783 fraudulent news. The Gossipcop and PolitiFact datasets are drawn from the entertainment and political sections of the Fake Newsnet repository. There are 244 actual newscast items and false things 135 in the PolitiFact training data, while there are 75 actual newscast items and false items 29 in the test data. Gossipcop has 10,010 training newscast pieces, from that 2,036 false newscast and actual news 7,974 make up the test data. However, the test data has 2,285 actual news items, 545 fraudulent news items. Because there are a lot of duplication problems in the Twitter dataset, we decided against doing tests on it; many posts lack images and a high percentage of tweets about single events, potentially leading to model overfitting.

Baseline Methods

To guarantee a fair and replicable assessment, it is important to meticulously choose baseline methods that consider the existence of pre-trained models or publicly accessible source code. Adherence to a standardized evaluation protocol that incorporates all three datasets for both training and testing is also important. The proposed MTLFND was evaluated against these selected methods, as given in Table 1.

Table 1: Baseline methods and their working

Model	Description
CAFÉ [18]	Creates an approach to identify misinformation in various forms, accounting for uncertainties by flexibly merging features from distinct modalities and their interrelationships.
CMC [40]	The approach entails commonly training two single-modal systems. Through the use of a cross-modal refinement objective function, CMC guides the single-modal networks with a soft target, making it easier to learn feature correlations between various modalities.
EANN [41]	The approach makes use of an event adverse networking topmost event model to overfit to a particular event.
SAFE [42]	Combines the relationship between textual and visual data into the classifier to identify fake news.
SPOTFAKE [43]	The approach utilize an article's linguistic and visual components. In particular, used language models (such as BERT) for learning features of text and pre-trained VGG-19 on the dataset ImageNet to acquire visual characteristics.

Implementation Details

In implementing the suggested multimodal fake news detection using a transfer learning framework, specific Python libraries were instrumental for efficient processing and model integration. Notably, the Hugging Face Transformers library was employed to access and fine-tune pre-trained language models tailored to the language of the data. The "RoBERTa-base-Chinese" model was utilized, while for English text data, the "RoBERTa-base-uncased" model was employed for English text data. Additionally, the framework utilized the TensorFlow library for attention-based post-processing techniques, enhancing the models' effectiveness in capturing intricate textual patterns. To standardize the input text length, preprocessing was conducted using TensorFlow to ensure a consistent length of 300 words per sample.

The MTLFND framework adopts specific pretraining models tailored to the language of the data. For Chinese data, we utilize the "RoBERTa-base-Chinese" pre-trained model, while for English data, we employ the "RoBERTa-base-uncased" model. Additionally, we apply attention-based post-processing to enhance the effectiveness of these models. The input text size is standardized to three hundred words to maintain sample uniformity.

It employs a ResNet-101 model that has been pre-trained on visual data, with the submitted image dimensions set to 224 × 224 pixels. To maintain alignment, the size of images inputted to MTLFND matches that of ResNet. Moreover, it generates summaries of texts exceeding 50 words using a summary generation model to meet input size specifications. The "ViTB/32" pre-trained model was used for this purpose.

During the training phase, ResNet is fine-tuned while ROBERTA and MTLFND weights remain fixed due to their complexity and the limited dataset size. The projection heads comprise two completely linked layers, each including 64 and 256 hidden units. Similarly, the classifier consists of hidden two fully connected layers of 64 and 2 units, correspondingly. The Adam optimizer with default settings and a batch size of 64 is used to train the model. The weight decay is set to 12, and the learning rate is set to 1 × 10-3. A training set of 50 epochs is chosen to reduce the possibility of overfitting, and the epoch with the best test accuracy is chosen.

RESULTS AND DISCUSSION

Experimental Results

The experimental results section thoroughly analyses and evaluates the proposed framework. Insights are offered to acquire from rigorous experimentation and validation conducted on diverse datasets encompassing both textual and visual modalities. The objective is to measure the MTLFND performance and efficacy in accurately identifying fake news through various linguistic and visual contexts. The measures for evaluation employed, the experimental setup utilized, and the comparative study with present state-of-the-art methods are discussed in detail. Through these investigations, this section aims to clarify the capabilities and MTLFND limitations, shedding light on its practical utility and prospective for applications of real-world in combating the proliferation of misinformation across multimedia platforms. The proposed method was applied to all three datasets, and their results are gathered in Table 2.

Thus, Table 2 presents a comparative analysis of the performance between suggested MTLFND and alternative methods across three datasets. Notably, the proposed approach, MTLFND, is the most accurate of these techniques. Furthermore, its success in identifying false news across various datasets is highlighted by the fact that its accuracy, F1 score and recall outperform the majority of the examined approaches.

Table 2: Comparative analysis of the performance

Dataset	Class Label	Parameter	CAFÉ [18]	CMC [40]	EANN [41]	SAFE[42]	SPOTFAKE [43]	Proposed MTLFND
Weibo	Fake News	Accuracy	0.840	0.908	0.827	0.762	0.892	**0.910**
		Precision	0.855	**0.940**	0.847	0.831	0.902	0.916
		Recall	0.830	0.869	0.812	0.724	**0.964**	0.901
		F1Score	0.842	0.899	0.829	0.774	**0.932**	0.908
	Real News	Precision	0.825	0.876	0.807	0.695	0.847	**0.916**
		Recall	0.851	**0.945**	0.843	0.811	0.656	0.901
		F1Score	0.837	0.907	0.825	0.748	0.739	**0.908**
Gossipcop	Fake News	Accuracy	0.867	0.893	0.864	0.838	0.858	**0.890**
		Precision	0.732	**0.826**	0.702	0.758	0.732	0.561
		Recall	0.490	0.657	0.518	0.558	0.372	**0.658**
		F1Score	0.587	0.692	0.594	0.643	0.494	0.754
	Real News	Precision	0.887	**0.920**	0.887	0.857	0.866	0.899
		Recall	0.957	0.963	0.956	0.937	0.962	**0.966**
		F1Score	0.921	**0.935**	0.920	0.895	0.914	0.928
PolitiFact	Fake News	Accuracy	0.882	0.895	0.870	0.905	0.851	**0.978**
		Precision	0.857	0.806	0.902	0.893	**0.953**	0.895
		Recall	**0.915**	0.862	0.825	0.908	0.733	0.895
		F1Score	0.885	0.833	0.862	0.901	0.828	**0.905**
	Real News	Precision	0.907	0.944	0.841	0.916	0.786	**0.960**
		Recall	0.844	0.920	0.912	0.901	0.964	**0.960**
		F1Score	0.876	0.932	0.875	0.890	0.866	**0.960**

Table 3: Average Performance of all model on all three datasets

Parameter	CAFÉ [18]	CMC[40]	EANN [41]	SAFE [42]	SPOTFAKE [43]	Proposed MTLFND
Average Accuracy	0.863	0.899	0.854	0.835	0.867	**0.926**
Average F1Score	0.824	0.866	0.818	0.809	0.795	**0.893**
Average Precision	0.696	**0.885**	0.831	0.825	0.847	0.858
Average Recall	0.672	0.708	0.811	0.806	0.775	**0.880**
Average Performance	0.763	0.839	0.828	0.818	0.821	**0.889**

Table 3 encapsulates the aggregated performance metrics, showcasing the average results attained by each method across all three datasets. The comprehensive summary offers a consolidated view of each model's efficacy in detecting fake news instances within diverse linguistic and visual contexts.

The table is offerings the mean accuracy, recall, precision and F1-score achieved by each method, providing valuable insights into their overall performance across the datasets. Through this comparative analysis, a deeper interpretation of the relative powers and weaknesses of the MTLFND framework and other examined approaches is gained, thereby facilitating informed decision-making and further refinement of fake news detection procedures.

Discussion of results

The MTLFND results on three well-known datasets, as Table 1 illustrates, reveal significant potential, with an average accuracy surpassing 91% on Weibo and exceeding 97% on PolitiFact. This suggests the proposed algorithm is dependable and resilient in detecting fake news across diverse multilingual and multidomain sources. Particularly noteworthy is the recall rate for real news on all three datasets, which exceeds 0.89, indicating that MTLFND is less likely to identify real news as phoney incorrectly.

Analyzing the correlation between cross-modal correspondence scores and the prevalence of bogus news on Gossipcop and Weibo datasets reveals that news articles are more likely to be genuine when similarity scores are higher. Furthermore, by categorizing news instances into bins based on their similarity scores, the corresponding dataset's average actual news rate is compared with the real news rates for each bin. These comparisons illustrate that on Weibo, the real news rate tends to increase with increasing ambiguity, while on Gossipcop, it initially declines before experiencing a surge. These statistical trends offer valuable insights for deep networks in identifying fake news.

The analysis of the results presented in Table 2 unveils intriguing perceptions into the performance of various fake news recognition models across multiple datasets. Among the examined methods, MTLFND emerges as a standout performer, boasting the highest average accuracy with 0.926 and the commendable F1-score with 0.893. This underscores the efficiency of the attention mechanism integrated into the MTLFND architecture in effectively capturing contextual dependencies within textual data. This underscores the efficiency of leveraging transfer learning and multimodal fusion procedures in enhancing the recognition capabilities of fake news across diverse linguistic and visual contexts. Additionally, MTLFND showcases competitive average precision and recall scores of 0.858 and 0.880, respectively, reaffirming its robustness in accurately identifying fake news instances. These outcomes underscore the prospective of MTLFND as a capable solution for combating the proliferation of misinformation across multimedia platforms, offering a compelling avenue for upcoming exploration and applications in fake news discovery.

Comparison with State-of-the-Art Approaches

Regarding accuracy, MTLFND performs better than all other examined algorithms across three datasets. Even while Spot on Weibo achieved somewhat higher recall, as shown in Figure 2, MTLFND notably attains the highest accuracy rates 91.0%, 89.0%, 97.0% correspondingly, exceeding the state-of-the-art method on the three real-world datasets. Furthermore, MTLFND routinely places first or second in all tests for accuracy, recall, and precision, demonstrating its efficacy.

The graphic compares the effectiveness of several cutting-edge techniques and the suggested MTLFND model for identifying false news using PolitiFact, Gossipcop, and Weibo datasets. Here's a breakdown of the results for each metric:

- Accuracy: The metric measures the proportion of correctly classified news articles (real or fake). MTLFND outperforms all compared methods on Gossipcop, Weibo and politiFact datasets.
- Precision: The metric measures the proportion of news articles identified as fake that are fake. CMC performs well on Weibo and Gossipcop but MTLFND stands second.

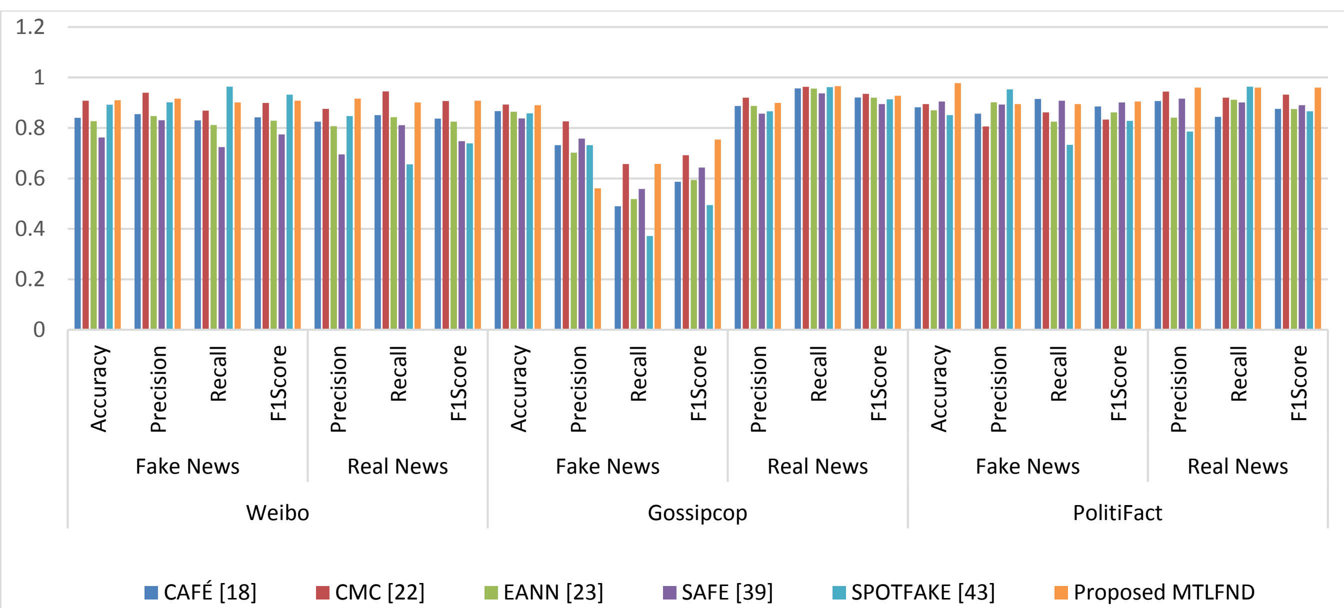


Figure 2: Comparison of Proposed MTLFND with other state-of-the-art method

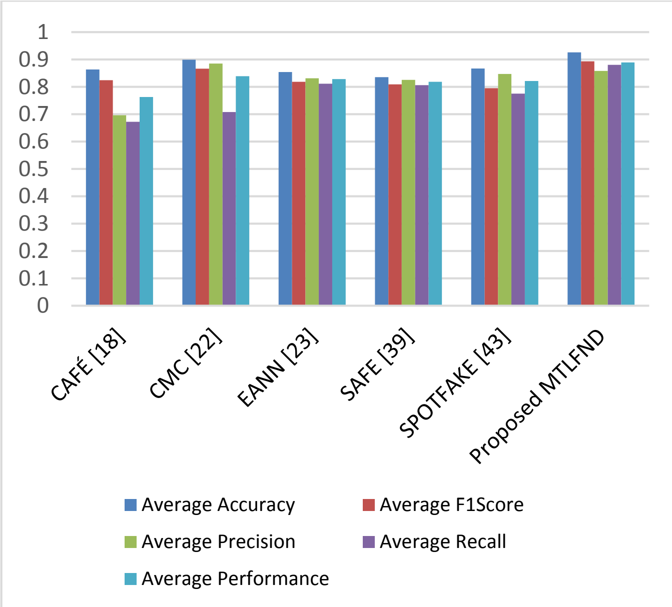


Figure 3: Average Performance of all state-of-the-art methods

- Recall: This statistic measures the percentage of real fake news stories that are accurately classified as fake. Although it performs lower than other algorithms on Gossipcop and PolitiFact, MTLFND has the most recall on the Weibo dataset.
- F1 Score: This metric, the harmonic mean of accuracy and recall, measures how well a system can distinguish phony news from actual news without misclassifying the former. MTLFND achieves the best F1 score on Weibo, Gossipcop and politiFact.

The MTLFND method performs well on false news detection, especially on the Gossipcop and politiFact datasets. It achieves high accuracy and F1 scores, as shown in Figure 3 while having good precision and recall.

The suggested system advances the state-of-the-art in false news identification by including several significant characteristics and innovative components. Leveraging the power of transfer learning, MTLFND integrates rich representations derived from text and image modalities using specialized encoders, including the encoder ResNet-based for images, the encoder RoBERTa for text. However, this reliance raises pertinent questions concerning the calculation of feature similarity and its subsequent impact on the decision-making process in fake news discovery. Moreover, while the prospective of pre-trained multimodal feature learning models holds promise, existing approaches have not fully exploited this potential. Conversely, the uniqueness of our suggested method, MTLFND, lies in the introduction of a comprehensive framework tailored to focus the limitations. By leveraging the power of transfer learning and multimodal feature extraction, our method improves the model's capability to mine crucial features for effectiveness of fake news finding.

CONCLUSION AND FUTURE SCOPE

In conclusion, a pioneering approach, MTLFND with transfer learning, is introduced for multimodal fake news finding, showcasing its efficacy in enhancing classification accuracy. Leveraging MTLFND, aligned features are extracted across text and image modalities, with modality-wise attention employed to dynamically assign weights and mitigate the impact of noisy features during fusion. Extensive experiments conducted on established FND datasets demonstrate that integrating MTLFND complements unimodal features, exceeding many current state-of-the-art techniques. Classification accuracy is significantly improved by MTLFND, which attains the highest accuracy of 91.0%, 89.0%, and 97.8%, correspondingly, surpassing the state-of-the-art method on Weibo, Gossipcop, and PolitiFact, three real-world data sets respectively. The binary nature of its output is acknowledged, providing limited insight into the rationale behind news classification. Future research endeavors aim to address this limitation by developing more interpretable fake news detection systems. These systems will not only classify news as "real" or "fake" but also provide a detailed understanding of the reasons behind the classification, enabling the identification of specific suspicious elements within news articles. Pursuing this avenue aims to contribute to the advancement of more transparent and interpretable fake news detection mechanisms in the evolving landscape of information authenticity.

CONFLICT OF INTEREST STATEMENT

It is declared by all authors that they have no conflicts of interest.

REFERENCES

1. X. Shen, M. Huang, Z. Hu, S. Cai, T. Zhou. Multimodal Fake News Detection with Contrastive Learning and Optimal Transport. *Front Comput Sci* **2024**, 6.
2. H. Allcott, M. Gentzkow. Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives* **2017**, 31 (2), 211–236.
3. X. Zhou, R. Zafarani. A Survey of Fake News. *ACM Comput Surv* **2021**, 53 (5), 1–40.
4. H. Guo, J. Cao, Y. Zhang, J. Guo, J. Li. Rumor Detection with Hierarchical Social Attention Network. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*; ACM, New York, NY, USA, **2018**; pp 943–951.
5. S. Tufchi, A. Yadav, T. Ahmed. A comprehensive survey of multimodal fake news detection techniques: advances, challenges, and opportunities. *Int J Multimed Inf Retr* **2023**, 12 (2), 28.
6. M. Nirav Shah, A. Ganatra. A systematic literature review and existing challenges toward fake news detection models. *Soc Netw Anal Min* **2022**, 12 (1), 168.
7. R. Kumari, N. Ashok, P.K. Agrawal, T. Ghosal, A. Ekbal. Identifying multimodal misinformation leveraging novelty detection and emotion recognition. *J Intell Inf Syst* **2023**, 61 (3), 673–694.
8. H.F. Villela, F. Corrêa, J.S. de A.N. Ribeiro, A. Rabelo, D.B.F. Carvalho. Fake news detection: a systematic literature review of machine learning algorithms and datasets. *Journal on Interactive Systems* **2023**, 14 (1), 47–58.
9. A. Galli, E. Masciari, V. Moscato, G. Sperlí. A comprehensive Benchmark for fake news detection. *J Intell Inf Syst* **2022**, 59 (1), 237–261.
10. M.R. Islam, S. Liu, X. Wang, G. Xu. Deep learning for misinformation detection on online social networks: a survey and new perspectives. *Soc Netw Anal Min* **2020**, 10 (1), 82.
11. K. Shu, D. Mahudeswaran, S. Wang, D. Lee, H. Liu. FakeNewsNet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media. *Big Data* **2020**, 8 (3), 171–188.
12. Y.L.Z.L.R.Y.Y.Z.D.W. Yifan Liu. Modality Interactive Mixture-of-Experts for Fake News Detection. *Mach Learn* **2025**, 2.

13. Y. Guo. A mutual attention based multimodal fusion for fake news detection on social network. *Applied Intelligence* **2023**, 53 (12), 15311–15320.
14. J. Jing, H. Wu, J. Sun, X. Fang, H. Zhang. Multimodal fake news detection via progressive fusion networks. *Inf Process Manag* **2023**, 60 (1), 103120.
15. B. Palani, S. Elango. CTrL-FND: content-based transfer learning approach for fake news detection on social media. *International Journal of System Assurance Engineering and Management* **2023**, 14 (3), 903–918.
16. Y. Jiang, X. Yu, Y. Wang, et al. Similarity-Aware Multimodal Prompt Learning for fake news detection. *Inf Sci (N Y)* **2023**, 647, 119446.
17. J. Xue, Y. Wang, Y. Tian, et al. Detecting fake news by exploring the consistency of multimodal data. *Inf Process Manag* **2021**, 58 (5), 102610.
18. Y. Chen, D. Li, P. Zhang, et al. Cross-modal Ambiguity Learning for Multimodal Fake News Detection. In *Proceedings of the ACM Web Conference 2022*; ACM, New York, NY, USA, **2022**; pp 2897–2905.
19. Y. Zhou, Y. Yang, Q. Ying, Z. Qian, X. Zhang. Multimodal Fake News Detection via CLIP-Guided Learning. In *2023 IEEE International Conference on Multimedia and Expo (ICME)*; IEEE, **2023**; pp 2825–2830.
20. E.F. Ayetiran, Ö. Özgöbek. An inter-modal attention-based deep learning framework using unified modality for multimodal fake news, hate speech and offensive language detection. *Inf Syst* **2024**, 123, 102378.
21. Y. Wu, P. Zhan, Y. Zhang, L. Wang, Z. Xu. Multimodal Fusion with Co-Attention Networks for Fake News Detection. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*; Association for Computational Linguistics, Stroudsburg, PA, USA, **2021**; pp 2560–2569.
22. S. Patel, S. Surati. Unmasking Fake News: Navigating the Landscape of Fake News Identification, Challenges and Issues. *SN Comput Sci* **2024**, 5 (7), 932.
23. J. Xue, Y. Wang, Y. Tian, et al. Detecting fake news by exploring the consistency of multimodal data. *Inf Process Manag* **2021**, 58 (5), 102610.
24. A. Vineela, A. Bhavani, B.V. Krishna, A.B. Sankar. An artful multimodal exploration in discerning fake news through text and image harmony. *Multimed Tools Appl* **2025**.
25. R.K. Kaliyar, A. Goswami, P. Narang. FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. *Multimed Tools Appl* **2021**, 80 (8), 11765–11788.
26. L. Zhang, X. Zhang, Z. Zhou, et al. Early Detection of Multimodal Fake News via Reinforced Propagation Path Generation. *IEEE Trans Knowl Data Eng* **2025**, 37 (2), 613–625.
27. S. Bayram. Image manipulation detection. *J Electron Imaging* **2006**, 15 (4), 041102.
28. S. Tufchi, A. Yadav, T. Ahmed. A comprehensive survey of multimodal fake news detection techniques: advances, challenges, and opportunities. *Int J Multimed Inf Retr* **2023**, 12 (2), 28.
29. I. Saha, S. Puthran. Fake News Detection: A Comprehensive Review and a Novel Framework. In *2024 OITS International Conference on Information Technology (OCIT)*; IEEE, **2024**; pp 463–469.
30. C. Comito, L. Caroprese, E. Zumpano. Multimodal fake news detection on social media: a survey of deep learning techniques. *Soc Netw Anal Min* **2023**, 13 (1), 101.
31. P. Yang, J. Ma, Y. Liu, M. Liu. Multi-modal transformer for fake news detection. *Mathematical Biosciences and Engineering* **2023**, 20 (8), 14699–14717.
32. F. Wu, S. Chen, G. Gao, Y. Ji, X.-Y. Jing. Balanced Multi-modal Learning with Hierarchical Fusion for Fake News Detection. *Pattern Recognit* **2025**, 164, 111485.
33. H. Lv, W. Yang, Y. Yin, et al. MDF-FND: A dynamic fusion model for multimodal fake news detection. *Knowl Based Syst* **2025**, 317, 113417.
34. A. Gandhi, K. Adhvaryu, S. Poria, E. Cambria, A. Hussain. Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions. *Information Fusion* **2023**, 91, 424–444.
35. A. Giachanou, G. Zhang, P. Rosso. Multimodal Fake News Detection with Textual, Visual and Semantic Information; **2020**; pp 30–38.
36. J. Li, X. Dai, R. Yan, C. Tang, Y. Li. Multi-anchor adaptive fusion and bi-focus attention for enhanced gait-based emotion recognition. *Sci Rep* **2025**, 15 (1), 14946.
37. J. Hua, X. Cui, X. Li, K. Tang, P. Zhu. Multimodal fake news detection through data augmentation-based contrastive learning. *Appl Soft Comput* **2023**, 136, 110125.
38. S. Qian, J. Wang, J. Hu, Q. Fang, C. Xu. Hierarchical Multi-modal Contextual Attention Network for Fake News Detection. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*; ACM, New York, NY, USA, **2021**; pp 153–162.
39. S. Patel, S. Surati. MTL-rtFND: Multimodal Transfer Learning for Real-Time Fake News Detection on Social Media; **2024**; pp 235–247.
40. R. Khamsehashari, V. Schmitt, T. Polzehl, S. Mohtaj, S. Möller. How Risky is Multimodal Fake News Detection? A Review of Cross-Modal Learning Approaches under EU AI Act Constrains. In *3rd Symposium on Security and Privacy in Speech Communication*; ISCA, ISCA, **2023**; pp 1–10.
41. X. Zhou, J. Wu, R. Zafarani. $$\mathsf {SAFE}$$: Similarity-Aware Multi-modal Fake News Detection; **2020**; pp 354–367.
42. J.W.R.Z. Xinyi Zhou. SAFE: Similarity-Aware Multi-Modal Fake News Detection. *Computer Science, Computation and Language* **2020**.
43. S. Singhal, R.R. Shah, T. Chakraborty, P. Kumaraguru, S. Satoh. SpotFake: A Multi-modal Framework for Fake News Detection. In *2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM)*; IEEE, **2019**; pp 39–47.