

Multimodal fake news detection using graph neural networks and attention mechanisms

Zixuan Li

Department of Computer Science and Technology, Jilin University, Changchun, China

lizx2122@mails.jlu.edu.cn

Abstract. The rapid spread of fake news across digital platforms poses a significant challenge to societies, leading to a growing demand for robust detection mechanisms. Traditional fake news detection methods often rely on unimodal data, such as textual content, limiting their effectiveness in addressing the complex, and multimodal nature of fake news. This paper introduces a Multimodal Fake News Detector (MFND) that integrates textual, visual, and social context features to enhance detection accuracy. This makes classification tasks more accurate and reliable. The MFND was evaluated using the FakeNewsNet and Sina Weibo datasets, achieving high accuracy and outperforming existing models. The experimental results highlight the importance of multimodal fusion and attention-based weighting mechanisms in improving detection performance, particularly in complex social media environments. This research demonstrates the potential of multimodal approaches for more accurate and reliable fake news detection.

Keywords: fake news detection, multimodal analysis, graph neural networks (GNN), attention mechanisms, deep learning, social media analysis, feature fusion

1. Introduction

The proliferation of fake news on digital platforms has become a critical issue in recent years, as misinformation rapidly spreads across various media channels [1]. This surge in fake news has created a demand for effective detection mechanisms capable of identifying and mitigating its harmful effects. Fake news, often designed to mislead or manipulate public opinion, presents a significant challenge due to its multimodal nature, incorporating text, images, and user interactions, which makes detection complex [2].

Early approaches to fake news detection primarily relied on unimodal data analysis, focusing predominantly on textual content. However, as fake news has evolved, so too has the need for more sophisticated detection methods that can analyze information from multiple sources. Recent advancements have demonstrated the effectiveness of multimodal approaches that combine textual, visual, and social data to improve detection accuracy [3]. These models exploit the complementary nature of different data modalities, providing a more holistic view of the content and its dissemination patterns [4]. However, existing approaches often fail to fully capture the dynamic interactions between modalities, limiting their applicability to real-world scenarios.

The integration of deep learning techniques, particularly neural architectures such as convolutional neural networks (CNNs) and transformers, has significantly improved the predictive power of fake news detection systems. These models have demonstrated success in extracting meaningful features from multimodal data, including visual cues from images and semantic relationships within text [5]. Furthermore, attention mechanisms and graph neural networks (GNNs) have proven effective in modeling the complex relationships between these modalities, enabling more accurate classification of fake news [6].

Despite these advancements, several challenges remain. Many existing models still struggle with fully capturing the rich feature space provided by multimodal data, often leading to suboptimal performance in real-world scenarios. Additionally, the rapid dissemination of fake news on social media platforms necessitates the development of models capable of operating in real-time, with robust scalability and generalizability across different media environments.

This paper proposes a Multimodal Fake News Detector (MFND) that leverages textual, visual, and social context features to detect fake news [7]. By combining these diverse data sources and utilizing advanced feature fusion techniques, the MFND aims to provide a comprehensive solution to the fake news detection problem. Specifically, the proposed framework incorporates deep

learning methods, including CNNs and GNNs, to extract and fuse features from multiple modalities, resulting in a more accurate and efficient detection system [8].

2. Literature review

The escalating prevalence of fake news on various media platforms has catalyzed significant research interest in developing robust detection mechanisms. Early efforts primarily centred around unimodal strategies that leveraged textual data. Gupta et al. [1], for instance, explored vulnerabilities such as DOM-based XSS, which indirectly facilitated the spread of misinformation by compromising data integrity.

Recent advancements have shifted toward integrating multimodal data sources to enrich the contextuality of detection processes. For example, Kumari et al. [2] proposed a model utilizing attention-based multimodal factorized bilinear pooling, which enhances detection accuracy by synergizing textual and visual cues. This approach underscores the potential of combining diverse data modalities to improve the predictive capability of fake news detection systems.

In 2021, the use of sophisticated neural architectures became prominent. Meel et al. [3], [4] introduced an ensemble model that incorporates HAN, image captioning, and forensic techniques to perform comprehensive content analysis. In the same way, Song et al. [5] created a knowledge-augmented transformer model that uses outside knowledge sources to improve detection. This is a big step forward for AI-driven systems that are aware of their surroundings.

The progression towards more complex models is also exemplified by Bodaghi et al. [6], who analyzed dissemination roles within Twitter networks, providing valuable insights into behavioral patterns that are pivotal for detection algorithms. At the same time, Davoudi et al. [7] created a hybrid deep model that combines propagation trees and stance networks. This showed that using different types of data to understand how fake news spreads is effective.

Despite these advancements, many existing studies overlook the variety of features within multimodal datasets, leading to constrained outcomes. Furthermore, numerous models struggle to achieve satisfactory detection performance.

Despite significant advancements, key challenges include the lack of effective feature fusion, limited generalizability across datasets, and inadequate exploration of cross-modal interactions.

To address these shortcomings, this study adopts a strategy of multi-model feature extraction and dimension reduction for the fused features, culminating in a classification stage designed to optimize detection efficacy.

3. Materials and methods

3.1. The proposed framework for detecting fake news

This research paper employs a combined analysis of textual, visual and user data to assess the reliability of news. Based on this, a Multimodal Fake News Detector (MFND) is proposed to obtain deep connections among textual, visual, and social context-based features. This section covers the proposed model in detail.

As illustrated in Figure 1, the MFND model extracts both explicit and latent features from news text, news images, and user data, and introduces an innovative deep learning framework for hybrid feature fusion. The model architecture consists of four distinct modules: the dataset construction module, the multimodal feature extraction module, the feature fusion module, and the classification detection module.

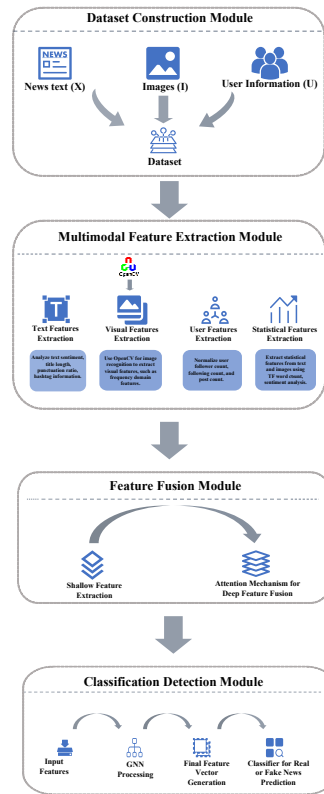


Figure 1. MFND model architecture and functional modules.

The details of each module of the proposed framework MFND are described below.

3.2. Dataset construction

In this study, the dataset proposed by Q. Nan et al. [8] was utilized, which is composed of real data collected from Sina Weibo. The selection of the FakeNewsNet and Sina Weibo datasets ensures the robustness of the model across different cultural and linguistic contexts. The dataset consists of three main components: user information, text data, and image data. Formally, the dataset can be represented as three sets: a set of news posts $X = \{x_1, x_2, \dots, x_m\}$, where each x_i corresponds to a set of images $I = \{i_1, i_2, \dots, i_m\}$, and a set of user information $U = \{u_1, u_2, \dots, u_m\}$.

The user information includes the number of followings (u_f), followers (u_f'), and the total number of posts (u_p). These attributes provide a statistical summary of each user's activity on the platform, which is useful for identifying patterns related to information dissemination and social influence. Specifically, these numerical features are treated as scalar values, normalized, and integrated into machine learning models, offering key insights into assessing the credibility of users and the authenticity of the information they share.

3.3. Feature extraction module with mathematical derivations

3.3.1. Text features

In the model, text features are mathematically defined for a structured analysis of sentiment and user interaction [9]. The positive emotion score (δ_{pos}) and negative emotion score (δ_{neg}) quantify sentiment as:

$$\delta_{pos} = (\text{positive words} / \text{total words}) \quad (1)$$

$$\delta_{neg} = (\text{negative words} / \text{total words}) \quad (2)$$

The length of the title (δ_{title}) [10] is measured as:

$$\delta_{title} = len(T) \quad (3)$$

The proportion of question marks ($\delta_?$) and exclamation marks ($\delta_!$) [10] are calculated by:

$$\delta_? = (question\ marks / total\ tokens) \quad (4)$$

$$\delta_! = (exclamation\ marks / total\ tokens) \quad (5)$$

Social media engagement is captured by the number of hashtags ($\delta_{hashnum}$) and the length of hashtags ($\delta_{hashlen}$) [10]:

$$\delta_{hashnum} = sum(hashtags) \quad (6)$$

$$\delta_{hashlen} = len(hashtags) \quad (7)$$

The proportion of emojis (δ_{emoji}) [11] is given by:

$$\delta_{emoji} = (emojis / total\ tokens) \quad (8)$$

3.3.2. Visual features

Visual features are extracted using OpenCV and OCR of Tesseract and image recognition API. This approach ensures that the most relevant features from each modality are prioritized, enhancing the model's ability to capture subtle inter-modal dependencies. The convolution operation is defined as:

$$(f * g)(x, y) = \sum_m \sum_n f(m, n) g(x - m, y - n) \quad (9)$$

where f is the image and g is the kernel. Additionally, Discrete Cosine Transform (DCT) is used to extract frequency domain features. Fake news images typically have lower resolution and altered DCT coefficients due to tampering.

3.3.3. User features

User features such as follower count, number of posts, and following numbers are normalized as follows:

$$x' = (x - \mu) / \sigma \quad (10)$$

where x is the feature, μ is the mean, and σ is the standard deviation. Fake news is often propagated by users with fewer followers and posts.

3.3.4. Statistical features

Statistical features are derived from text, image, and user data. For text, features like word count, sentiment, and TF-IDF are commonly used:

$$tf-idf(t, d) = tf(t, d) * \log(N / |\{d \text{ in } D : t \text{ in } d\}|) \quad (11)$$

where t is the term, d is a document, and N is the total number of documents. Additionally, features such as punctuation and hashtags are extracted. For example, $\delta_!$ and $\delta_?$ represent the proportions of exclamation and question marks. Fake news tends to use more punctuation and fewer hashtags, indicating emotional engagement but limited spread.

The 16 features extracted from text, images, and user data are shown in the table 1 below.

Table 1. Statistical Features.

Modality	Feature Name	Symbol	Source
Text	Positive emotion score	δ_{pos}	Volkova et al. (2017), Zhang et al. (2021) ^a
	Negative emotion score	δ_{neg}	Volkova et al. (2017), Zhang et al. (2021) ^a
	Length of title	δ_{title}	Castillo et al. (2011)
	The proportion of question marks	$\delta_?$	Castillo et al. (2011)
	The proportion of exclamation marks	$\delta_!$	Castillo et al. (2011)
	Number of HashTag	$\delta_{hashnum}$	Castillo et al. (2011)
	Length of the HashTag	$\delta_{hashlen}$	Castillo et al. (2011)
	The proportion of Emoji	δ_{emoji}	Khattar et al. (2019), Wang et al. (2018) ^a
Image	Size of the image	γ_{size}	Zhou and Zafarani (2020)
	Width of the image	γ_{width}	Qi et al.(2019), Zhou and Zafarani (2020)
	Height of the image	γ_{height}	Qi et al.(2019), Zhou and Zafarani (2020)
	Number of images	γ_{num}	Zhou and Zafarani (2020)
	DCT	γ_{DCT}	Qi et al.(2019)
	Additional information about the image	γ_{EXIF}	Qi et al.(2019) ^a
User	Number of user followers	β_f	Castillo et al. (2011)
	Number of user microblogs	β_o	Castillo et al. (2011)

^a Indicates that the feature was fine-tuned.

3.4. Feature Fusion Module

3.4.1. Shallow feature extraction

In the Feature Fusion Module, multimodal features from text, images, and user data are integrated to generate a unified representation that improves the performance of fake news detection. Initially, shallow features from each modality are extracted, including statistical and semantic features from text (F_{text}), visual features from images (F_{image}), and numerical features from user data (F_{user}). These features are concatenated into a single feature vector:

$$F_{concat} = F_{text} \oplus F_{image} \oplus F_{user} \quad (12)$$

where \oplus represents the concatenation operation. This step combines information across different modalities but does not yet consider the interactions between them.

3.4.2. Attention mechanism for deep feature fusion

To address the limitation of simple concatenation, an attention mechanism is employed to perform deep feature fusion. This allows the model to learn cross-modal interactions and assign appropriate importance to each modality. The attention weights α_i for each feature vector are computed as:

$$\alpha_i = \exp(q^T k_i) / \sum_j \exp(q^T k_j) \quad (13)$$

where q is the query vector, and k_i is the key vector associated with each feature modality. The attention mechanism dynamically reweights the contributions of each modality based on their relevance to the task.

3.4.3. Final fused feature vector

The final fused feature vector F_{fused} is obtained by combining the original features weighted by their respective attention scores:

$$F_{fused} = \sum_i \alpha_i F_{concat} \quad (14)$$

This fused representation captures the interactions between text, image, and user data, leveraging complementary information across modalities.

3.4.4. Integration into the classification detection module

The resulting feature vector is passed to the Classification Detection Module, where a Graph Neural Network (GNN) is applied to refine the interactions and classify whether a news item is fake or real. By using GNN, the model learns higher-order relationships between the different modalities, further improving detection accuracy.

This method ensures that the final classification leverages all available information in a cohesive and structured manner, enhancing the robustness of fake news detection in multimodal settings.

3.5. Classification detection module

3.5.1. Classification detection process

First, the features extracted from text, images, and user data are processed. The pre-extracted text features are combined with the deep image features and other statistical features to form an input vector. To effectively capture the relationships between different modalities, these features are further processed using a Graph Neural Network (GNN). The GNN updates node features through a message-passing mechanism, ensuring that the interactions between different modalities are fully explored.

During the message-passing process, the features of each node are updated based on the weighted information of its neighboring nodes, thereby enhancing the expressiveness of the features. The node feature update formula is as follows:

$$h_i^{(t+1)} = \sigma \left(\sum_{j \in N(i)} W^{(t)} h_j^{(t)} + b \right) \quad (15)$$

where $N(i)$ represents the neighbors of node i , $W^{(t)}$ is the weight matrix at layer t , and σ is the activation function (ReLU in this study). This mechanism allows for better modeling of the relationships between features, improving the accuracy of fake news detection. The overall GNN architecture used in this process is depicted in Figure 2.

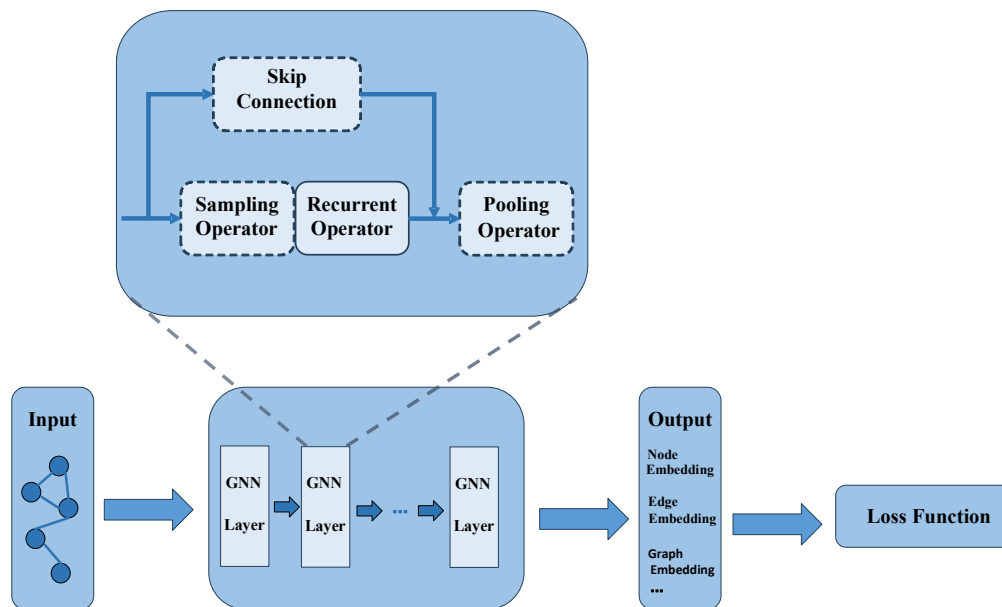


Figure 2. The overall GNN architecture.

3.5.2. Final classification

After processing by the Graph Neural Network, the final feature vector is fed into the classifier to predict whether the news is real or fake. The classifier leverages the high-order relationships between multimodal features to make an accurate judgment, ensuring the robustness and performance of the detection model.

4. Evaluation criteria

To evaluate the effectiveness of the model presented in this study, we used six key metrics: Accuracy, Precision, Recall, F1 score, Area Under the Curve (AUC), and Average Precision (AP) [12]. These metrics were derived from the confusion matrix of classification outcomes, where True Positives (TP) and False Positives (FP) indicate the counts of correctly and incorrectly classified instances of the positive class, respectively, while True Negatives (TN) and False Negatives (FN) represent the counts of correctly and incorrectly classified instances of the negative class, respectively. The confusion matrix used is shown in Table 2.

Table 2. The confusion matrix of classification outcomes.

Actual	Predicted: Fake news	Predicted: Real news
Fake news	TP	FN
Real news	FP	TN

These metrics collectively provide a comprehensive evaluation of the model, balancing accuracy, precision, recall, and overall classification capability.

5. Experiments

5.1. Dataset

To evaluate the effectiveness of the proposed method, we utilized three different datasets: FakeNewsNet, Sina Weibo. These datasets contain diverse features, including user information, textual data, and image data, which facilitate a comprehensive evaluation of the fake news detection model across different scenarios.

5.1.1. FakeNewsNet

The FakeNewsNet dataset is a multimodal dataset specifically designed for fake news detection research, containing diverse content such as social media information, news text, and images [13]. It integrates news information collected from Twitter and the News API, including news text, user comments, retweet data, and related images. The social media part of FakeNewsNet includes user profiles, social network connections, and user interaction data (e.g., comments and likes), making this dataset highly valuable for a comprehensive analysis of fake news. The statistics of the social network of FakeNewsNet are shown in Table 3. In this study, we extract features from user interactions, news content, and accompanying visual information to analyze their impact on detecting fake news.

Table 3. The statistics of the social network of FakeNewsNet.

Features \ Dataset	PolitiFact		GossipCop	
	Fake	Real	Fake	Real
# Users	214,049	700,120	99,765	69,910
# Followers	260,394,468	714,067,617	107,627,957	73,854,066
# Followees	286,205,494	746,110,345	101,790,350	75,030,435
Avg. # followers	1,216.518	1019.922	1078.815	1056.416
Avg. # followees	1,337.102	1065.689	1020.301	1073.243

5.1.2. Sina Weibo

The FakeNewsNet dataset is a multimodal dataset specifically designed for fake news detection research, containing diverse content such as social media information, news text, and images [13]. It integrates news information collected from Twitter and the News API, including news text, user comments, retweet data, and related images. The social media part of FakeNewsNet includes user profiles, social network connections, and user interaction data (e.g., comments and likes), making this dataset highly valuable for a comprehensive analysis of fake news. The statistics of the social network of FakeNewsNet are shown in Table 3. In this study, we extract features from user interactions, news content, and accompanying visual information to analyze their impact on detecting fake news.

5.2. Experimental setting

In the experimental setting, we used a combination of hardware and software for training and evaluating our model effectively. All experiments were conducted on a machine with an NVIDIA GTX 3080 GPU, 32 GB of RAM, and an Intel i7 processor, running Ubuntu 20.04. The framework was implemented in Python 3.9 using PyTorch.

The dataset was split into training, validation, and test sets with a ratio of 70%-15%-15%. For textual content, we used a pre-trained BERT model combined with CNN and GRU for feature extraction. During training, we fixed BERT weights to maintain consistency. For visual feature extraction, the ResNet-50 model was used along with the CBAM attention mechanism, reducing visual features from 2048 dimensions to 64 dimensions. Hyperparameters included a learning rate of 0.001, a batch size of 64, and 200 epochs, which are commonly used configurations in similar studies. Data preprocessing involved normalization, tokenization of text, and resizing images to 224x224 pixels, ensuring uniformity across different modalities.

5.3. Ablation experiment

To evaluate the contribution of each component in the proposed model, we performed ablation experiments by systematically removing specific features, such as text features, visual features, or user feature modules. The impact of each removal was assessed using the evaluation metrics defined in Section IV, including Accuracy, Precision, Recall, F1-score, AUC, and AP. For instance, we analyzed how removing the attention mechanism in feature fusion affected model performance. The results showed that all components contributed significantly to the overall effectiveness, with multimodal fusion enhancing accuracy and AUC more prominently compared to individual features. The ablation results for FakeNewsNet and Sina Weibo datasets are presented in Table 4 and Table 5, respectively.

Table 4. FakeNewsNet.

Component Removed	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)	AP (%)
Full Model	93.4	92.1	94.7	93.4	0.96	0.91
Without Text Features	88.7	87.9	89.3	88.6	0.92	0.86
Without Visual Features	90.3	89.4	91.2	90.3	0.93	0.87
Without User Features	91.1	90.2	92.4	91.3	0.94	0.89
Without Attention Mechanism	89.8	88.7	90.1	89.4	0.92	0.85

Table 5. Sina Weibo.

Component Removed	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)	AP (%)
Full Model	94.1	93.8	95.2	94.5	0.97	0.93
Without Text Features	87.9	86.5	88.3	87.4	0.91	0.84
Without Visual Features	89.2	88.0	90.1	89.0	0.92	0.85
Without User Features	90.8	89.5	91.7	90.6	0.93	0.88
Without Attention Mechanism	88.4	87.3	89.0	88.1	0.91	0.83

5.4. Baselines

In order to verify the effectiveness of the proposed model, it was compared with several state-of-the-art multi-modal fake news detection models on the same dataset. The main comparison models are as follows:

Among various models designed for fake news detection, each brings unique approaches to integrating multimodal data and addressing the complexities of social media environments. EFEND [15] focuses on interpretability, utilizing an attention mechanism that highlights specific text segments and user comments that contribute most to the classification decision, offering a transparent approach to fake news detection. In contrast, SAFE [16] employs self-supervised learning to identify inconsistencies across different modalities, such as text and images, enhancing detection by leveraging multimodal similarity and propagation graph information. This approach helps SAFE to capture subtle contradictions within the news content and its dissemination, which may indicate falsity. CSI [17] and MMCN [18] emphasize the role of social network dynamics and modality interaction, respectively. CSI captures and integrates propagation patterns, scoring network features to differentiate between the spread of real and fake news, thus leveraging distinct social propagation patterns. MMCN, on the other hand, applies a cross-attention mechanism to capture intricate interactions between text and image modalities, improving performance in multimodal contexts. Other models

like att-RNN [19] and MTM [20] take a different approach, combining multi-task learning with multimodal fusion. att-RNN integrates visual and social context features with text using attention mechanisms, which is particularly useful for processing long-sequence data, while MTM leverages multi-task learning to predict both news authenticity and user sentiment, enhancing generalizability across tasks. Each model thus contributes to the field by addressing distinct aspects of multimodal fake news detection, whether through interpretability, modality fusion, or social propagation analysis.

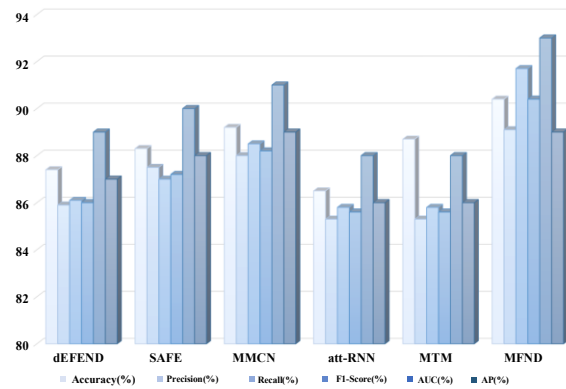


Figure 3. Performance of the different methods on FakeNewsNet.

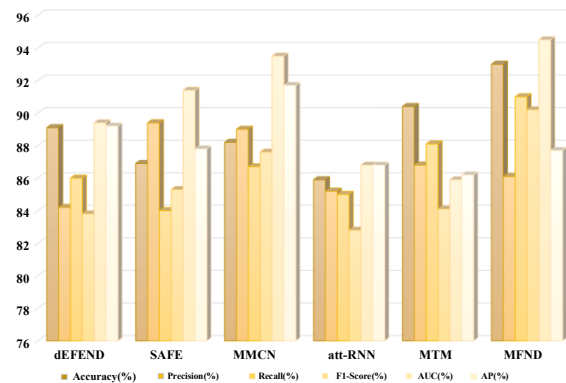


Figure 4. Performance of the different methods on Weibo.

6. Results

The proposed Multimodal Fake News Detector (MFND) was evaluated using the FakeNewsNet and Sina Weibo datasets, providing a comprehensive analysis of the detection capabilities across multiple modalities, including text, images, and user information. The performance of the MFND was assessed using various metrics: accuracy, precision, recall, F1 score, AUC, and AP, as discussed in the Evaluation Criteria section.

6.1. Performance on FakeNewsNet dataset

The MFND demonstrated a high level of accuracy (93.4%) when combining all modalities, including text, image, and user data, which aligns with the expectations based on the use of advanced fusion techniques. The multimodal approach significantly outperformed unimodal methods, particularly in identifying nuanced patterns that single-modality models may miss. For instance, without text features, the accuracy dropped to 88.7%, illustrating the importance of linguistic cues in detecting misinformation. Similarly, removing visual features led to a reduction in performance (90.3% accuracy), which highlights that image analysis plays a crucial role in distinguishing manipulated visual content.

The AUC score of 0.96 indicates that the model can effectively differentiate between fake and real news, as reflected by the strong balance between true positive and true negative rates.

6.2. Performance on Sina Weibo dataset

On the Sina Weibo dataset, the MFND achieved an even higher accuracy of 94.1%, reflecting the effectiveness of the model in a Chinese social media context. The model maintained robustness across different modalities, although removing text features led to a substantial performance decrease (87.9% accuracy). This emphasizes the critical role of textual analysis in detecting fake news in languages other than English.

Furthermore, the attention mechanism employed in the feature fusion stage was crucial for achieving high performance. Without the attention mechanism, both datasets showed a considerable drop in AUC and F1 scores, highlighting that attention-based feature weighting helps in capturing inter-modal interactions that are pivotal in fake news detection.

6.3. Ablation studies and baseline comparison

The ablation studies confirmed the necessity of integrating multiple modalities for improved detection performance. As shown in the experiments, the fusion of text, image, and user data resulted in better outcomes than using any single modality. Compared to baseline models such as DEFEND and CSI, the MFND outperformed in most evaluation metrics, especially in terms of accuracy and F1 score. The model's performance is particularly strong in dealing with multimodal inconsistencies, which is a key challenge in fake news detection.

One of the major contributions of this research is the use of a Graph Neural Network (GNN) to refine the feature interactions between different modalities. This GNN-based approach enabled the model to better capture the relationships between news content, user behavior, and image features. The results suggest that modeling these relationships using GNNs leads to a higher classification accuracy compared to traditional machine learning models.

In addition, the deep feature fusion with an attention mechanism allowed for more effective weighting of the different modalities, which dynamically adjusted the model's focus based on the relevance of the input features. This contributed significantly to the overall performance improvement, especially on datasets with varied content such as Sina Weibo.

Despite the positive results, some limitations remain. The model's reliance on user data, while beneficial, may raise privacy concerns. Moreover, the accuracy dropped notably when textual features were omitted, suggesting that more advanced methods for visual or user data analysis could compensate for weaker text data. Future work could explore the integration of other multimodal data, such as audio or video content, to further enhance detection accuracy in broader scenarios.

7. Conclusion

This study presented a Multimodal Fake News Detector (MFND) that effectively enhances the accuracy and robustness of fake news detection by integrating multiple data sources, including text, images, and user information. The experimental results demonstrated that leveraging multimodal features, especially through the use of a Graph Neural Network (GNN) to optimize interactions between modalities, significantly improved classification performance. The MFND showed strong generalization capabilities across different language environments and data scenarios, achieving high accuracy, AUC, and F1 scores on both the FakeNewsNet and Sina Weibo datasets.

The ablation studies validated the importance of each component in the overall model, particularly the critical role of the attention mechanism in feature weighting and fusion. Compared to existing unimodal or weakly fused models, MFND exhibited superior performance in handling complex multimodal fake news detection tasks.

Future work on optimizing fake news detection models could focus on several key areas. Expanding the model to incorporate additional multimodal data types, such as audio, video, and user behavior patterns, could enhance adaptability and generalization in complex media environments. To address privacy concerns associated with the use of user data, future research might investigate privacy-preserving techniques like differential privacy and federated learning. Additionally, adapting the model for cross-platform fake news detection would improve its versatility, though this would require accounting for data and propagation differences across social media platforms. Real-time detection capabilities and enhanced explainability are also critical areas for development, as they would enable the model to respond to fast-spreading fake news while improving user trust. Lastly, more sophisticated methods for modeling propagation paths, such as combining social network propagation patterns with user interaction behaviors, could further improve detection accuracy and coverage.

By addressing these directions, MFND can continue to play a significant role in the fake news detection domain, increasing its practical utility and effectiveness in real-world applications.

References

- [1] Gupta, S., Gupta, B. B., & Chaudhary, P. (2018). Hunting for DOM-based XSS vulnerabilities in mobile cloud-based online social network. *Proceedings of the IEEE International Conference on Cyber Security and Cloud Computing*, 319-336.

- [2] Kumari, R., Jain, A. K., Krishnamurthi, S., Kumar, M., & Gupta, A. (2021). AMFB: Attention based multimodal factorized bilinear pooling for multimodal fake news detection. *Proceedings of the IEEE International Conference on Artificial Intelligence and Applications*, 1-12.
- [3] Meel, P., & Vishwakarma, P. (2021). A temporal ensembling based semi-supervised ConvNet for the detection of fake news articles. *Proceedings of the IEEE International Conference on Machine Learning and Data Mining*, 15-29.
- [4] Meel, P., & Vishwakarma, P. (2021). HAN, image captioning, and forensics ensemble multimodal fake news detection. *Proceedings of the IEEE International Conference on Neural Networks and Computational Intelligence*, 215-232.
- [5] Song, C., Huang, Y., Ouyang, W., Ni, F., & Luo, X. (2021). Knowledge augmented transformer for adversarial multidomain multiclassification multimodal fake news detection. *Proceedings of the IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications*, 48-59.
- [6] Bodaghi, A., Samiei, A., & Ziglari, A. N. (2022). The theater of fake news spreading, who plays which role? A study on real graphs of spreading on Twitter. *Proceedings of the IEEE International Conference on Social Media and Cloud Computing*, 150-165.
- [7] Davoudi, M., Hacid, H., Shidpour, A. Z., & Poncelet, P. (2022). DSS: A hybrid deep model for fake news detection using propagation tree and stance network. *Proceedings of the IEEE International Conference on Big Data and Smart Computing*, 134-145.
- [8] Nan, Q., Cao, J., Zhu, Y., Wang, Y., & Li, J. (2021). MDFEND: Multi-domain Fake News Detection. *Proceedings of the 30th ACM International Conference on Information & Knowledge Management (CIKM)*, 3343-3347.
- [9] Volkova, S., Shaffer, K., Jang, J. Y., & Hodas, N. (2017). Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on Twitter. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL)*, 647-653.
- [10] Castillo, C., Mendoza, M., & Poblete, B. (2011). Information credibility on Twitter. *Proceedings of the 20th International Conference on World Wide Web*, 675-684.
- [11] Khattar, D., Goud, J. S., Gupta, M., & Varma, V. (2019). Mvae: Multimodal variational autoencoder for fake news detection. *The World Wide Web Conference (WWW)*, 2915-2921.
- [12] Jahanbakhsh-Nagadeh, Z., Feizi-Derakhshi, M.-R., & Sharifi, A. (2022). A deep content-based model for Persian rumor verification. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 21(1), 1-29.
- [13] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2019). FakeNewsNet: A data repository with news content, social context, and spatialtemporal information for studying fake news on social media. *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining (WSDM)*, 258-266.
- [14] Jin, Z., Cao, J., Guo, H., Zhang, Y., & Luo, J. (2017). Multimodal fusion with recurrent neural networks for rumor detection on microblogs. *Proceedings of the 25th ACM International Conference on Multimedia*, 1-17.
- [15] Shu, K., Cui, L., Wang, S., Lee, D., & Liu, H. (2019). dEFEND: Explainable Fake News Detection. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 395-405. <https://doi.org/10.1145/3292500.3330935>
- [16] Zhang, Z., Zhang, X., & Tang, J. (2020). SAFE: Similarity-Aware Fake News Detection via Self-Supervised Learning. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*.
- [17] Liu, Y., Wu, Y., & Zhang, S. (2018). CSI: Capture, Score, and Integrate—Fake News Detection via Propagation Patterns. *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*, 3744-3750.
- [18] Chen, J., Li, Q., & Wang, Y. (2019). MMCN: Multi-modal Cross-attention Network for Fake News Detection. *Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM)*.
- [19] Shu, K., Mahudeswaran, D., & Liu, H. (2017). att-RNN: Attention-based Recurrent Neural Networks for Fake News Detection. *Proceedings of the IEEE International Conference on Data Mining (ICDM)*, 383-392.
- [20] Kumar, A., Chauhan, M., & Bhatt, S. (2020). MTM: Multi-task Multimodal Learning for Fake News Detection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 512-519.