ARTICLE

A News Media Bias and Factuality Profiling Framework Assisted by Modeling Correlation

Qi Wang[1], Chenxin Li[1,*], Chichen Lin[2], Weijian Fan[3], Shuang Feng[1] and Yuanzhong Wang[4]

[1]School of Computer and Cyber Sciences, Communication University of China, Beijing, 100024, China

[2]State Key Laboratory of Media Convergence and Communication, Communication University of China, Beijing, 100024, China

[3]School of Data Science and Intelligent Media, Communication University of China, Beijing, 100024, China

[4]Beijing 797 Audio Co., Ltd., Beijing, 100016, China

*Corresponding Author: Chenxin Li. Email: cuc_lcx@cuc.edu.cn

ABSTRACT

News media profiling is helpful in preventing the spread of fake news at the source and maintaining a good media and news ecosystem. Most previous works only extract features and evaluate media from one dimension independently, ignoring the interconnections between different aspects. This paper proposes a novel news media bias and factuality profiling framework assisted by correlated features. This framework models the relationship and interaction between media bias and factuality, utilizing this relationship to assist in the prediction of profiling results. Our approach extracts features independently while aligning and fusing them through recursive convolution and attention mechanisms, thus harnessing multi-scale interactive information across different dimensions and levels. This method improves the effectiveness of news media evaluation. Experimental results indicate that our proposed framework significantly outperforms existing methods, achieving the best performance in Accuracy and F_1 score, improving by at least 1% compared to other methods. This paper further analyzes and discusses based on the experimental results.

KEYWORDS

News media profiling; factuality; bias; correlated features

1 Introduction

As media formats diversify and the speed of news dissemination accelerates, the proliferation of subjective tendencies and misleading information presents significant challenges to the public's ability to accurately understand facts and form independent judgments. Researchers have undertaken various initiatives to address this issue, such as stance detection, and fake news detection [1]. However, the detection speed of these methods is far from enough for the current quantity of news and is not competent for the credibility identification of news in the early stage of dissemination. Consequently, many researchers have shifted their focus towards analyzing news organizations.

As the primary medium for disseminating information and guiding public opinion, news media reflects the overall quality of published content. The formation strategy of the present news aggregation platform is mainly to satisfy the needs of users, which may cause users to be trapped in the situation of information cocoon and filtering foam [2]. When users consistently engage with media that exhibits extreme bias and low factual accuracy, their perspectives can become increasingly narrow. This gradual erosion of awareness regarding mainstream ideologies hampers the development of a healthy informational environment. Profiling news media can improve the quality of news reports from the media sources and promote the sustainable development of transparency and fairness of information communication [3]. Most methodologies tend to evaluate credibility, factuality, or bias independently, treating these aspects as isolated tasks without thoroughly exploring the interrelationships among their features. Most methodologies tend to evaluate credibility, factuality, or bias independently, treating these aspects as isolated tasks without thoroughly exploring the interrelationships among their features.

There are more explorations in finer-grained rumor detection methods, many of which adopt a multi-task learning framework to jointly model stance detection and rumor verification [4–8]. These two tasks can mutually promote each other. Some approaches leverage stance information to detect rumors based on tree structures [9,10]. In media profiling tasks, corresponding features can be extracted by embedding models and pre-trained language models [11–13]. Additionally, some researchers have introduced additional knowledge distillation frameworks to distill auxiliary knowledge [14]. A well-established connection exists between factuality and bias, so it is logical to jointly model these two aspects [15]. In addition, the assessment methodology and guidelines from the public media evaluation website Media Bias/Fact Check (MBFC) indicate that its rating of factuality considers bias. Conversely, its evaluation of bias also takes into account factual integrity as well as the use of credible and verifiable sources. A multi-task framework approach [15] employs multi-task training to concurrently model bias and factual relationships. Although this approach acknowledges their interconnection, it does not explicitly incorporate a relationship extraction module within its structural framework. Graph-based methods represent news content in various graph forms [15–17], utilizing graph neural networks to extract pertinent features for classification purposes. Partisan bias often leads individuals to accept news that aligns with their ideological beliefs while disregarding the authenticity of news that contradicts those beliefs [18], thereby diminishing the accuracy in identifying fake news. Numerous fake news detection methodologies [19] have already integrated stance modeling with authenticity assessments. The classification of stance regarding rumors is deemed a crucial step in predicting their authenticity since stance influences judgment accuracy concerning rumors. Wei et al. [20] proposed a hierarchical multi-task learning framework aimed at jointly predicting rumor stance and accuracy on Twitter.

In order to further investigate media profiling tasks and enhance their performance, this paper discusses the relationship between the factuality and political bias of media sources, as well as the design of a framework informed by this connection. Based on a review of previous studies, it is posited that there exists a correlation between the factuality and bias of news media. Media outlets exhibiting extreme political bias tend to disseminate content that diverges from established facts, resulting in lower levels of factuality. This research employs label distribution heat maps and calculates correlation coefficients to validate this relationship. Furthermore, this paper proposes a novel media profiling framework aimed at fully leveraging this correlation while maintaining the independence of evaluation tasks. Specifically, feature extractors are constructed based on pre-trained language models to capture both overall bias and factuality within media articles. The aggregate expression of the articles represents the features of the media, so our focus lies in modeling relationships among

articles rather than those contained within individual articles. Feature alignment and fusion modules are used to capture implicit correlations, employ recursive convolution and attention mechanisms as strategies, and additionally utilize global and local feature enhancement modules composed of pooling layer and multi-layer perceptron to further capture multi-scale information. Ultimately, the fused features are input into a classifier to yield evaluation results. Through experimental validation, this method demonstrates significant improvements in evaluation performance across multiple indicators. Moreover, this paper analyzes the effectiveness of the proposed framework through exploratory experiments including ablation studies and feature visualization pertaining to model structure. In summary, our main contributions are as follows:

- Firstly, this research explores and discusses the potential correlation between media factuality and political bias, which can complement each other in media profiling tasks.
- Secondly, a new news media profiling framework assisted by correlated features is proposed, which comprehensively utilizes dynamic attention fusion and high-order sequence interaction to enhance the model's evaluation of news media sources. This framework incorporates a recursive gating mechanism and depthwise separable convolution, as well as global and local feature enhancement modules to model this correlation.
- Thirdly, this paper designs and conducts extensive and rigorous experiments on the proposed framework, comparing multiple methods, performing ablation studies, and analyzing from a visualization perspective, which also verify the effectiveness of our approach.
- This paper is structured as follows. The introduction outlines the background and articulates the research questions. Subsequently, the related work section provides a comprehensive overview of existing literature in the field. The methodology section details the research design and model architecture employed in this study. The experiments section presents an overview of the dataset, experimental setup, analysis methods, key findings of the research, and discusses their implications. Finally, the conclusion summarizes the research outcomes, acknowledges limitations, and proposes directions for our future investigation.

2 Related Work

As the amount and variety of information increase, information evaluation has been studied at different levels, including (i) Claim-level, (ii) Article-level, (iii) User-level and (iv) Medium-level [12]. At the claim-level, it mainly adopts the form of fact-checking to check the supporting relationship between each part of evidence and the original text, and uses the retrieved information as evidence to check whether the statement conforms to the fact. At the article level, many studies have revealed the great potential of incorporating stance-aware knowledge into rumor verification to boost verification performance [9]. For example, rumors can be debunked by cross-checking the stance conveyed by their related posts, which also depends on the attributes of the rumor [10]. These two closely related tasks are jointly learned. Yang et al. [10] transformed two multi-class problems into multiple MIL-based binary classification problems, using a hierarchical attention mechanism to aggregate. MATNN designed a structural representation called RC-Trees to combine the two tasks [7]. Chen et al. jointly formulated a multi-stage classification task and proposed a multigraph neural network framework to construct hierarchical heterogeneity to promote cross-topic feature propagation [8]. Luo et al. used a partition filter network to explicitly model rumor and stance features and shared interaction features [5]. SSRI-Net designed a position rumor interactive network to fully integrate [21]. Credibility and stance-aware recursive tree (CSATree) validated for short retweets [9]. Ma et al. proposed a stance-aware multi-task model DSMM to determine the authenticity of statements based on dialogue graphs [6]. Krieger et al. employed domain-adaptive pre-training approach to enhance the detection of language

bias in news articles, validate it on multiple pre-trained language models, and solve the challenge of identifying bias in different linguistic contexts [22]. Some multi-modal methods use complementary attention mechanisms to fuse multiple types of information [23,24]. Liu et al. collected articles written by different political ideological media for pre-training, which is more advantageous in ideological prediction and position detection tasks [25].

The news media profiling tasks that this paper mainly studies belong to the medium level. By profiling news media, we can more effectively control the quality of news content from the sources and ensure the sustainability of the news environment. Furthermore, the evaluation results from the sources can be used as a prior to assist evaluation tasks at other levels. As shown in Table 1, it is the summary of the existing medium-level media profiling methods of four categories, including their contributions and limitations. The details are described below.

Baly et al. [11,12] first discussed the importance of predicting the reports' authenticity and news media bias, and modeled it as a classification task. They published their collected data and the labels mainly came from MBFC. They used support vector machines (SVM), Glove, and BERT to obtain rich and diverse features from various sources, e.g., articles, Twitter, and Facebook, and then conducted experiments on a large number of news media. They also explored the impact of news media content and readers on media profiling tasks. Although their work has advanced knowledge in media profiling, it primarily addresses the tasks in isolation, neglecting the interconnection of accuracy evaluations related to bias and factuality. Esteves et al. [26] proposed a model that can automatically extract source reputation clues and calculate the credibility factor of media sources. While their efforts provide insights into correlated tasks, they do not directly leverage these correlations for enhanced prediction accuracy. Besides, a multi-task ordinal regression framework jointly modeled the two issues of predicting media factuality and bias [15]. It is a straightforward way to exploit correlation. The authors also used some auxiliary tasks to improve the performance of the model. MiBeMC proposed by Fan et al. [1] constructed a similarity module, an interaction module, and an aggregation module to fully simulate human evaluation behavior. This approach effectively captures some aspects of media evaluation but may still overlook comprehensive feature interactions.

There are also some graph-based methods, researchers use graphs to construct connections between sources and news. Gruppi et al. [16] formalized the content-sharing behavior of news sources into a network and performed a random walk over the content-sharing network to detect the veracity of news sources. Because people tend to communicate with others with similar interests, GREENER [15] constructed a media connection graph based on audience overlap and used graph neural networks for news media profiling. Mehta et al. [17] constructed heterogeneous graphs to capture the interaction of social information and news content and iteratively extended the graph using inference operators based on similarity. Relationship graph convolutional network (R-GCN) is used for node representation. These focus on structural relationships rather than the nuanced interaction of contextual features such as bias and factuality. Burdisso et al. [27] proposed a reinforcement learning framework that models problems as estimates of reliability based on Markov decision processes. They constructed a larger dataset based on news published by Common Crawl and merged ground-truth labels from different sources. Hounsel et al. [28] explored techniques for detecting disinformation websites by analyzing various infrastructure characteristics, such as domain registration details and hosting services, to distinguish them from legitimate sites. Carragher et al. [13] considered search engine optimization as an effective attribute for predicting website reliability, using domains to construct graph connections but still lack the context-sensitive integration of various media attributes. Yang et al. [29] found that the ratings of ChatGPT are correlated with those of human experts,

indicating that large language models can also be an affordable reference for credibility ratings in fact-checking applications. Although large language models possess a more comprehensive and extensive capacity for semantic understanding and domain knowledge, they also exhibit certain illusions and deficiencies inherent to the model itself. These limitations can result in media assessments that are not necessarily more reliable.

Table 1: A summary of the existing medium-level media profiling methods

Method	Contributions	Limitations
Traditional machine learning-based methods [11,26]	Relatively interpretable and effective feature extraction from diverse sources of universities for media evaluation.	The limitation of feature engineering and poor adaptability.
Deep learning models-based methods [1,12,29]	Complex structural design and the use of models with inherent knowledge can extract deeper hidden information, which is friendly to downstream tasks.	The independent construction evaluation task ignores the correlation between articles and association attributes between media profiling tasks.
Graph-based methods [13,15,17,27]	Innovatively using graphs to imitate and fit news dissemination, making it more dynamic.	Focus more on structural information and ignore some context features.
Infrastructure character-based methods [28]	Innovative media evaluation from an engineering perspective.	Poor reliability, ignoring the content published by news media sources.

According to existing literature, the primary objective of most methodologies is to enhance the accuracy of specific media profiling tasks. They use different structures, information, and forms to construct models, but ignore the correlation between the two tasks. Although some researchers [15] have discussed this relationship and conducted experiments in the form of multi-task modeling, they have not directly extracted correlation to assist tasks. To address these limitations, this paper proposes a method to directly introduce relevant features into the framework. By leveraging convolutional and attention mechanisms, our approach harmonizes and enriches the data representation, thereby enhancing the robustness of media profiling tasks. This model not only capitalizes on the inherent complementarity of bias and factuality features but also demonstrates a significant enhancement in profiling reliability compared to existing methods.

3 Method

To better exploit the correlation between media bias and factuality, and use it to assist in improving media profiling performance, this study proposes a simple and effective framework for fusing and enhancing the correlation in this paper.

3.1 Correlation between Bias and Factuality

The political bias attribute of news media is related to factuality. For instance, media outlets with extreme-left or extreme-right orientations often exhibit propagandistic tendencies, whereas neutral media are generally more truthful and trustworthy. Biased processing of news information can impact detection accuracy with social media endorsement cues [30]. Biases in news perceptions and partisan leanings also influence the fact-checking results [31], so stance detection is considered a core task within fact-checking. Factuality and bias have some commonalities as they exert negative influences on the public by delivering information that deviates from the truth [15]. As the overall representation of all the news it releases, the media should also have this connection. This correlation can be clearly reflected in the label distribution heat maps of the dataset in Fig. 1. Fig. 1a,b respectively illustrates the correspondence between the labels of factuality and bias of the publicly available datasets EMNLP-2018 [11] and ACL-2020 [12]. The horizontal and vertical axes represent the corresponding label categories, with darker colors indicating a higher number of media with that label. The dark part of the heat map basically presents a V-shape, indicating the correlation between the two labels.
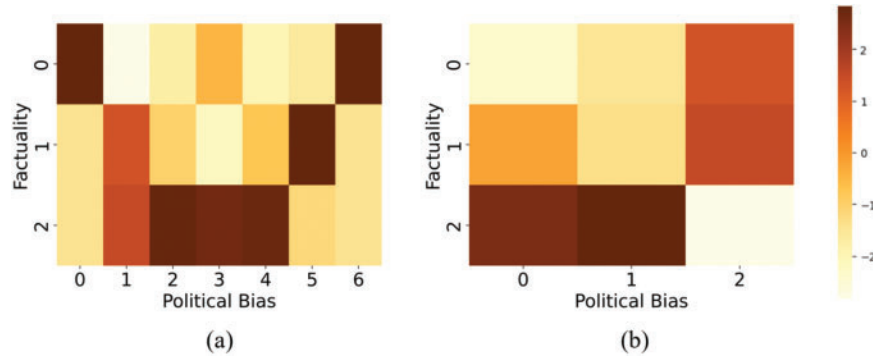

Figure 1: Label distribution heat maps of EMNLP-2018 and ACL-2020. Subfigure (a) depicts the relationship between the number of political bias and factuality labels in EMNLP-2018, and subfigure (b) presents the relationship between the number of bias and fact labels in ACL-2020

Pearson correlation coefficient, Spearman's rank correlation coefficient, and significance level p-values are also calculated in this paper. The p-value is determined by the observed correlation and the sample size, it proves that the correlation is significant, not accidental. Table 2 shows the indicator results of EMNLP-2018 and ACL-2020. The p-values of both correlation coefficients are small, and the absolute values of the correlation coefficients are generally above 0.4, confirming this connection. Therefore, we have reason to believe that the factuality and bias features of news media are related, so this study designs a feature fusion structure to integrate them to improve the effectiveness of news media profiling tasks.

Table 2: Correlation coefficient and p-value of datasets

Dataset	Pearson		Spearman	
	Correlation	p-value	Correlation	p-value
EMNLP-2018	-0.487	1.472e-64	-0.455	1.464e-55
ACL-2020	-0.563	7.827e-73	-0.587	7.805e-81

3.2 Proposed Method

This section introduces our framework in detail. Section 3.2.1 presents the overall framework structure, and Sections 3.2.2 and 3.2.3 introduce the implementation details of the main structures in our framework.

3.2.1 News Media Bias and Factuality Profiling Framework Assisted by Correlated Features

The architecture of our framework is depicted in Fig. 2. Firstly, our approach fine-tunes the pre-trained language model RoBERTa with two linear layers as our feature extractors. These extractors are utilized to derive bias and factuality representations of news media sources based on their published news. The extracted features are added together and sent to the main fusion structure in our framework. This structure extracts and aligns shared multi-scale information between features through a sophisticated interactive structure design, encompassing both local and global features. The classifier predicts the results based on the associated auxiliary features and consists of two linear layers. To improve the robustness and generalization ability of the model, we incorporate adversarial training techniques during the training process. By introducing adversarial samples, the model learns more essential and stable features of data and prevents over-fitting. Specifically, this study employs the adversarial training method known as the Fast Gradient Sign Method.

Since addition and concatenation are difficult to understand in the context, soft selection can take into account all input information and make weighted selections based on weights. The main idea of our framework is to utilize multi-scale contextual features, integrated with factuality and bias metrics, to derive attention weights that model the final relevant features [32]. Local features capture the association information between the local semantics of the text, while global features provide the general trends and relationships. Initially, the framework uses the recursive gated convolution method, which is implemented by combining pointwise convolution and depthwise convolution. This research designates this approach as Depthwise Separable Convolution with Recursive Gating (DSCRG). This structure leverages the strengths of convolutional neural networks and attention mechanisms to capture long-term dependencies and contextual interactions effectively. Subsequently, the feature enhancement module further extracts local and global features to enable the model to better learn content at different levels. This multi-level learning process enhances the overall feature representation capability of the model.

The attention weight calculation formula in the framework is as follows:

$$\omega = \sigma \left(L(X) + G(X) \right) = \sigma \left((L(X_b + X_f) + G(X_b + X_f) \right), \tag{1}$$

where $X \in \mathbb{R}^{S \times H}$, X_b and X_f are bias features and factuality features extracted by the Bias Extractor and Factuality Extractor, S denotes the length of input sequence, H denotes the hidden size. The input of extractors is news published by media sources, and the output is $X_b \in \mathbb{R}^{S \times H}$ and $X_f \in \mathbb{R}^{S \times H}$. Our approach adds the input features X_b and X_f, and feeds them into the aggregator and these two branches to obtain local and global features. $L(X)$ and $G(X)$ represent local context features and global context features, respectively. σ is a sigmoid function. $L(X)$ and $G(X)$ are as follows:

$$L(X) = E(R(X)), \tag{2}$$
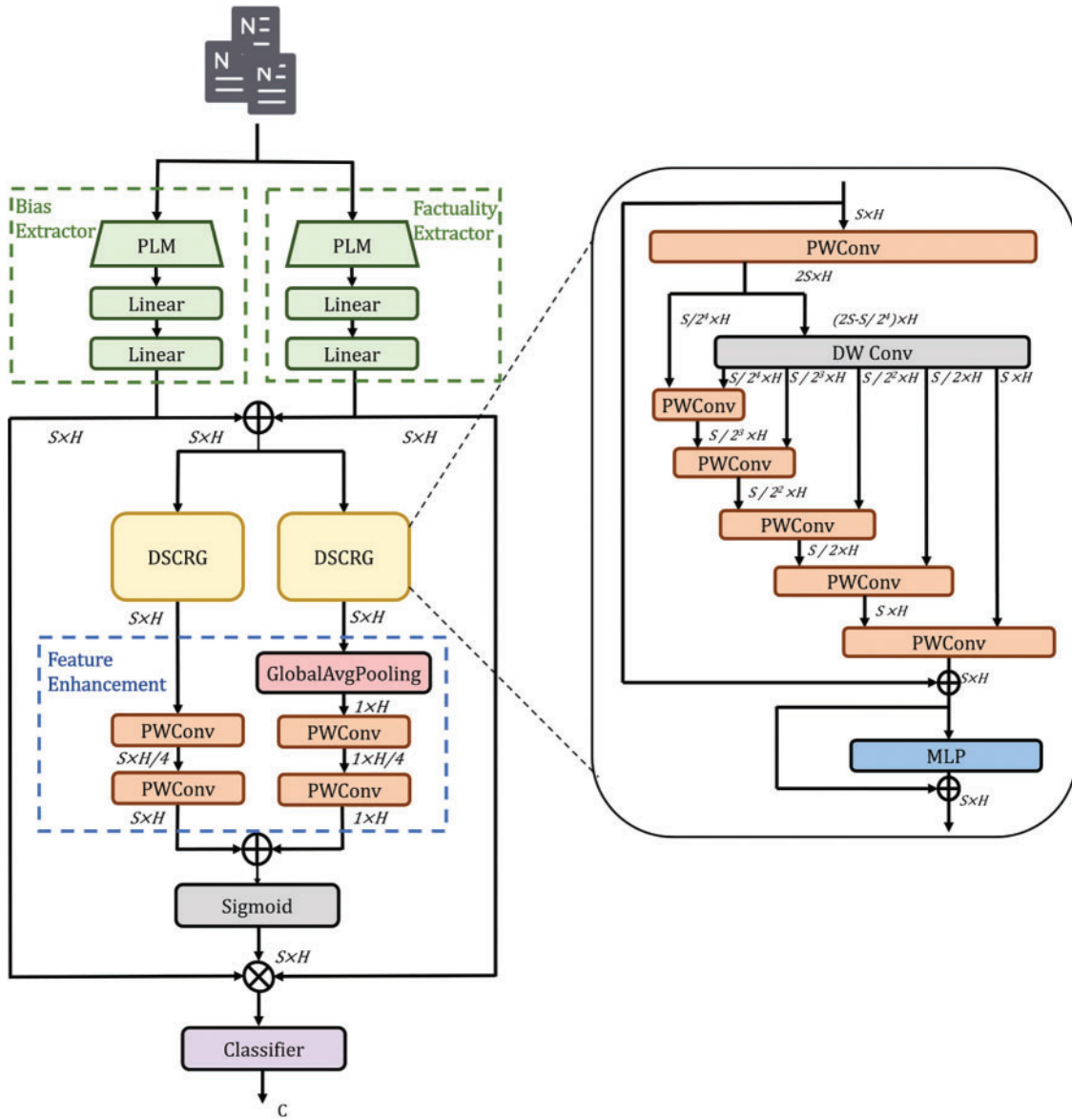
$$G(X) = E(P(R(X))). \tag{3}$$

Figure 2: Architecture of our proposed framework

R is DSCRG, P is global average pooling, E is feature enhancement module. More specific structural descriptions and analysis are shown in Sections 3.2.2 and 3.2.3.

$$C = Classifier\ (\omega \cdot X_b + (1 - \omega) \cdot X_f). \tag{4}$$

Classifier is composed of two linear layers.

3.2.2 DSCRG: Depthwise Seperable Convolution with Recursive Gating

Our approach uses DSCRG to obtain long-term high-order sequence interaction features to capture implicit connections between articles. Inspired by [33–36], pointwise convolution and depthwise

convolution can capture features of different scales and abstract levels in the text. Recursive and gating mechanisms are used to improve modeling capabilities and the model's ability to understand the semantics of text while keeping it as lightweight as possible.

As shown in Fig. 2, multi-order convolutions are used in DSCRG, which concatenates deep convolution from coarse to fine, ensuring that it does not introduce too much computational overhead and can complete high-order interactions. In addition, our proposed method also uses convolutional layers with large convolution kernels, making it more receptive and easier to capture long-term dependencies. DSCRG can better model complex interactive information and enable the correlation between bias and factuality features to be more deeply explored.

The input features $X_{in} \in \mathbb{R}^{S \times H}$ in DSCRG are projected through the projection layer ϕ to achieve high-order interaction, expanded to twice the original in the sequence length dimension and divided into two parts: d and p, which are used to feed depthwise convolution and pointwise convolution, which this paper refers as *DWConv* and *PWConv*, for step-by-step recursion. The depthwise seperable convolution with recursive gating is calculated as follows:

$$\left[d^{S_0 \times H}, \; p^{S_1 \times H} \right] = \phi \; (X_{in}) \in \mathbb{R}^{2S \times H}, \tag{5}$$

$$\left[p_0^{S_0 \times H}, \; p_1^{S_1 \times H}, \; \ldots, p_4^{S_4 \times H} \right] = DWConv \; (p), \tag{6}$$

$$d_k^{S_k \times H} = PWConv \; (p) \odot p_k^{S_k \times H}, k = 0, \; 1, \; \ldots, 4. \tag{7}$$

Our method splits the features after depthwise convolution to recursively complete the gated interaction with p. In this paper, it is split into 5, S_k is equivalent to $S/2^{n-k-1}$.

$$X_{out} = \varphi \; (d_4^{S_4 \times H}) + X_{in}, \tag{8}$$

$$Y = MLP \; (LN \; (X_{out})) + X_{out}. \tag{9}$$

Then the features are projected through a projection layer φ, with the dimensions remaining unchanged. Finally, our approach performs residual learning on the X_{out} obtained from the last layer, normalizes it through the layer normalization LN, and sends it to MLP with one hidden layer to get Y. In order to better enhance the subsequent global and local features, DSCRGs used by different bypasses in this framework do not share parameters.

3.2.3 Local and Global Context Feature Enhancement

After extracting features with rich interactive information from the above structure, in order to further align and model information, our approach extracts multi-scale local and global context features. The representation information obtained by DSCRG is used as local features, while global features are derived using global average pooling. Global average pooling summarizes spatial information [37], captures global context, and represents it in a manner that highlights the common features across the entire data. Then our method employs convolution to aggregate context at the embedding layer dimension. By using convolution with a kernel size of 1, our proposed method can integrate the input feature information to generate outputs of the same length, preserving the position information of the input sequence. Additionally, our framework uses bottleneck structure calculation to extract more refined and accurate feature representations. The calculation formula of the feature enhancement

module E is as follows:

$$Y_{local} = BN \ (Conv2 \ (\theta \ (BN \ (Conv1 \ (Y))))), \tag{10}$$

$$Y_{global} = BN \ (Conv2 \ (\theta \ (BN \ (Conv1 \ (P \ (Y)))))), \tag{11}$$

where $Conv1$ and $Conv2$ are pointwise convolution layers, BN is batch normalization, θ is a RELU activation function, and P is the global average pooling. Y represents the output features in DSCRG. $Conv1(Y) \in \mathbb{R}^{S \times H/4}$, it returns to its original size after $Conv2$.

$$\omega = sigmoid \ (Y_{local} + Y_{local}). \tag{12}$$

After merging the local and global context features, the weight coefficient is calculated by the Sigmoid function, and the classification result is calculated by Eq. (4).

4 Experiments

4.1 Datasets

To demonstrate the effectiveness of our proposed framework, this paper uses a public dataset commonly used in news media profiling ACL-2020, which contains lists of media domains and factuality (high, mixed, low) and bias (left, center, right) labels from MBFC. Due to the limitations of open source available datasets, this research collected more media sources labels and news data to verify the effectiveness of the proposed method on more data. The labels are collected from the same label source website according to the method of Baly et al., and the labels and data are processed in the same way. The labels of 1359 news media sources are from MBFC, removed sources with left-center and right-center labels and merged specified labels, kept the factuality and bias labels in a 3-category. 250,713 news published on these media sources are also collected, including titles and text content. Table 3 provides statistics of our data.

Table 3: Statistics of our collected data

Factuality			Bias		
Label	Sources	News	Label	Sources	News
Low	299	19,380	Left	277	28,448
Mixed	390	29,634	Center	723	28,336
High	670	49,130	Right	364	41,360

This study performed a classification task predictions of media factuality and bias, i.e., bias-based news media factuality profiling and factuality-based news media bias profiling, respectively. When predicting bias in news media, samples from all factuality categories were utilized. This comprehensive approach enables the model to effectively extract both the bias and factual style of the published news content under each media outlet, while exploring potential commonalities between these two attributes. Factuality prediction is also handled in the same way. Comparative and ablation experiments were conducted on two datasets, with results reported in terms of accuracy and F_1 score.

4.2 Experimental Setup

Multiple methods are used as benchmarks for these two profiling tasks. They include some state-of-the-art methods and other baseline methods:

- Majority class.
- Baly et al.: This method uses NELA features [38,39] from the article, averaged GloVe word embeddings to embed representations of Wikipedia pages, metadata from media Twitter profiles and URL structure features to train an SVM for classification.
- Baly et al.: Fine-tuning BERT to obtain article features, it aggregates YouTube, Facebook, and Twitter information, and trains an SVM classifier.
- DA-RoBERTa, 2022: The Transformer-based model RoBERTa is trained to solve the domain-specific problems.
- POLITICS, 2022: A model continued pre-training on RoBERTa based on collected diverse ideological leanings and language usage data.
- Mehta Best, 2022: Reasoning operators are used to expand the graph according to node similarities in the embedding space. It uses R-GCN to obtain node representations.
- MiBeMC, 2023: A structure that mimics human verification behavior in the network, including an informative feature extractor and a similarity module to obtain additional information, as well as an aggregation module.

Hyperparameters are set as follows: epoch is 100, the dropout rate is 0.4, the learning rate is set to $5e-4$, and an early stopping mechanism is set to prevent overfitting. Due to the length of the news, our method only retains titles and truncates the input text. Random seed is set to 42, and our code runs on one single NVIDIA RTX A4090 GPU.

4.3 Results and Discussions

The efficacy and progressiveness of our framework are demonstrated through comparative experiments with other methods, ablation studies, and visualization techniques. Compared to other methods, our framework effectively extracts and processes the complexity of media content by leveraging pre-trained language models and distinguishing feature extraction from relationship modeling. This separation preserves the independence of profiling tasks and more accurately captures implicit associations between articles, rather than relying solely on the internal structure or content of the articles. By using feature alignment and fusion modules, combined with recursive convolution and attention mechanisms, hidden correlations are effectively captured, and multi-scale features of information are more accurately captured. The global and local feature enhancement modules further allow information to be captured at multiple scales, providing a more comprehensive reflection of media characteristics and thereby improving overall performance.

4.3.1 Factuality Prediction

This study conducted news media factuality profiling experiments on ACL-2020 and our collected data. The results are shown in Table 4, our approach is compared with various methods, including a simple method such as directly classifying into the most numerous label type, traditional classical methods such as [11,12], method based on language model pre-training [24], and newer methods such as using graphs to simulate interaction processes [17] and designing complex networks and interaction module combinations [1].

Table 4: News media factuality profiling results

Model	ACL-2020		Our data	
	ACC	F_1 score	ACC	F_1 score
Majority class	52.43	22.93	49.30	22.01
Baly et al. [11]	66.45	61.08	69.53	65.91
Baly et al. [12]	71.52	67.25	67.84	64.24
Mehta BEST, 2022 [17]	72.55	66.89	–	–
DA-RoBERTa, 2022 [24]	71.06	64.98	73.08	71.35
MiBeMC, 2023 [1]	71.84	69.97	75.90	73.02
Ours	77.45	70.98	77.77	76.22

Since Baly et al. did not release the articles and social media data they used, Mehta et al. re-crawled data and followed their settings to replicate it [17]. This paper also reports the accuracy and F_1 score of our proposed framework as in previous work. Experimental results prove the advantages of introducing our framework. Our method achieved the best results in ACL-2020 with 86.91% and 86.13% in accuracy and F_1 score, respectively, an improvement of 4.9% in accuracy over the Mehta best and a 1.01% improvement in F_1 score over MiBeMC. This paper also shows the experimental results of existing methods on our own collected data. Since we only get news published by news media, the methods of Baly et al. only use the features extracted from news titles and content. Our framework also performed better than DA-RoBERTa, indicating that our model has acquired more evaluation knowledge compared to merely using pre-training models. Our method is also more stable compared to graph-based methods like Metha BEST. Compared with the MiBeMC with the best performance, our method improves 1.87% and 3.2% on accuracy and F_1 score, respectively, achieving performance of 77.77% and 76.22%. Evaluating media using internally stable relationships is more reliable than behavior-based methods that simulate human judgment. These results underscore the advantage of our framework in using correlated features to assist in media profiling. The high-order complex structure designed to extract the correlation between biased and factual content is indeed beneficial in assisting factual evaluations and enhancing the performance of factual tasks.

4.3.2 Bias Prediction

Table 5 shows our results on the bias media profiling task on ACL-2020 and collected data. Our proposed framework also shows great performance, outperforming other methods on the ACL-2020 dataset with an accuracy of 86.91% and an F_1 score of 86.1%. Compared to MiBeMC, it improved by 0.95% and 1.19%, respectively. Our method achieves significant improvement over other methods on the collected data, with an accuracy of 80.74% and an F_1 score of 80.09%, which is at least 1.9% higher in accuracy and 4.93% higher in F_1 score than other methods. In terms of bias assessment, our model has significantly improved in accuracy and F_1 compared with the trained method using political bias information [24,25].

The experimental results also demonstrate that introducing factual and bias correlation features can assist in improving the effectiveness of news media bias profiling. Our approach is more sensitive to the construction of associations to complete the assessment than the method of directly injecting information with pre-training. This method of mining correlation features is feasible and effective and

proves that our proposed framework structure, which utilizes complex convolutional networks and attention mechanisms, can extract representation information from the overall content of the article and model the correlations between contents.

Table 5: News media bias profiling results

Model	ACL-2020		Our Data	
	ACC	F_1 score	ACC	F_1 score
Majority class	40.39	19.18	53.20	23.15
Baly et al. [11]	73.61	72.90	74.54	70.65
Baly et al. [12]	85.29	84.77	71.38	66.83
DA-RoBERTa, 2022 [24]	83.72	82,98	76.73	70.12
POLITICS, 2022 [25]	86.06	84.65	77.15	76.45
MiBeMC, 2023 [1]	85.96	84.94	78.84	75.16
Ours	86.91	86.13	80.74	80.09

4.3.3 Visualization Analysis

Visualization offers a more intuitive and interpretable means to demonstrate the effectiveness of our proposed framework in media profiling tasks. This paper clusters and visualizes the feature vectors extracted from the model to show the differentiation and representativeness of the model, assessing whether the model can accurately reflect the differences of different categories. Our method uses uniform manual approximation and Projection (UMAP) [40] on the ACL-2020 data to reduce the dimensionality of the linear layer output features before softmax classification and further visualize them. As shown in Fig. 3, Fig. 3a,c is the visual clustering results of the factuality and bias labels of our proposed framework. Fig. 3b,d is the visualization results of the method that only uses a single feature extractor without introducing our auxiliary framework. Compared with Fig. 3b, the clustering effect of Fig. 3a is significantly better, with the higher distinction between different categories, forming independent clusters without overlap and mixing, and the data points of the same category are more concentrated. The distances between categories in Fig. 3c,d are obvious, and the data within the classes are relatively clustered. However, the number of data with the ground-truth label 0 or 2 that are misclassified in the predicted category with label 1 in Fig. 3c is less than that in Fig. 3d. Table 6 displays the results of the Silhouette Coefficient (SC) and Calinski-Harabasz Index (CH) to measure the rationality and quality of clustering. It demonstrates that the clustering results of (a) and (c) are better than those of (b) and (d).

These results also demonstrate that after introducing correlated features to assist the model in profiling media sources, the evaluation performance is significantly improved, which not only proves the effectiveness of our approach but also the usability of our proposed framework.
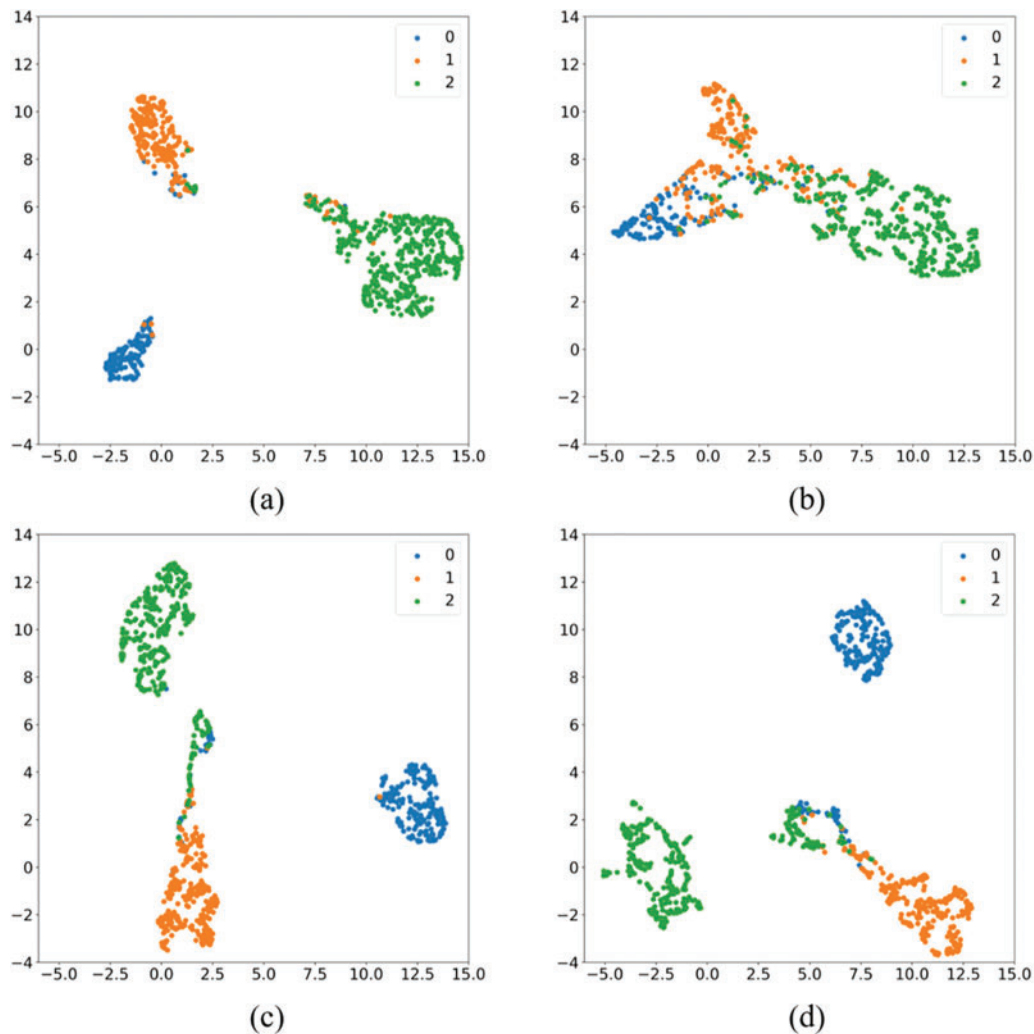
Figure 3: Visualization of clustering results based on UMAP. (a) displays the visual clustering result of the factuality labels of our proposed framework; (b) depicts the visualization effects of the factuality labels of the method that relies solely on a single feature extractor without introducing our proposed framework; (c) displays the visual clustering result of the bias labels of our proposed framework; (d) shows the visualization effects of the bias labels of the method that uses a single feature extractor without introducing our auxiliary framework

Table 6: Clustering evaluation index results

Fig. 3	SC	CH
(a)	0.654	2150.208
(b)	0.382	998.379

(Continued)

Table 6 (continued)

Fig. 3	SC	CH
(c)	0.672	2216.051
(d)	0.640	1918.00

4.4 Ablation Study

In Table 7, this research evaluates the results of removing each module of the main structure (i.e., DSCRG and feature enhancement modules) of our proposed framework and conduct detailed ablation experiments to verify the availability of the module combination we proposed. This study reports the experimental results of removing the main structure, DSCRG, and feature enhancement modules separately. In the media profiling tasks of bias and factuality, the results demonstrate that these methods perform worse than the methods using the entire framework in accuracy and F_1 score. Moreover, retaining any module in the structure does not show obvious advantages, but has a better performance when combined. This shows that our structure has a complementary effect. By extracting information in multiple dimensions and levels, it can assist in extracting and modeling correlated features, resulting in a better evaluation performance.

Table 7: Ablation study of our proposed framework

Model	ACL-2020		Our data	
	ACC	F_1 score	ACC	F_1 score
Factuality				
Ours$_{wo\ Enhance}$	69.77	70.16	75.30	70.80
Ours$_{wo\ DSCRG}$	68.14	67.67	76.65	73.08
Ours$_{wo\ Framework}$	70.01	69.56	77.41	72.55
Bias				
Ours$_{wo\ Enhance}$	74.91	72.96	77.44	77.42
Ours$_{wo\ DSCRG}$	76.41	69.62	77.96	78.01
Ours$_{wo\ Framework}$	76.58	69.35	77.17	74.43

In addition to removing different structures to verify the effectiveness of different modules, our approach also adjusted the order of modules in the framework and demonstrated through experiments that the module architecture we currently use can better model information connections. Our proposed method applied DSCRG after the feature enhancement module and conducted experiments on the data we collected. Due to the limitation of dimensional changes, our method only moved DSCRG after the local feature extraction module in the experiment, and the order of extracting global features remained, called Ours$_{after}$. This paper also conducted experiments by placing the DSCRG module both before and after the feature enhancement module, called Ours$_{before\ +\ after}$. The experimental results are shown in the Table 8. The performance of the framework decreases significantly after the adjustment of the module structure order. It is speculated that this drop in performance is due to the architectural

adjustments preventing the model from learning effective information, thus hindering its ability to accurately evaluate media sources. Specifically, placing the DSCRG after the enhancement module likely results in the extracted local information failing to cover the correlation between bias and factuality content. This leads to inaccurate correlation features that cannot be effectively utilized for subsequent assistance. Moreover, setting DSCRG both before and after the enhancement module would overly complicate the network structure, increasing the difficulty of modeling and making it challenging for the model to align the learned relational features. Therefore, the structure of our current framework is reasonable, which can ensure obtaining preliminary relevant content first and then strengthening it locally and globally to ultimately obtain features that are helpful for bias and factual evaluation of news media.

Table 8: Ablation study of different framework architectures

Method	ACC	F_1 score
Factuality		
$\text{Ours}_{\text{after}}$	75.12	73.85
$\text{Ours}_{\text{before + after}}$	74.63	66.67
Bias		
$\text{Ours}_{\text{after}}$	78.62	77.63
$\text{Ours}_{\text{before + after}}$	76.98	76.05

5 Conclusion

This study explores the intricate relationship between bias and factuality within the context of news media profiling tasks. Recognizing the critical interplay between these elements, this study proposes an innovative profiling framework that leverages their correlation to enhance the accuracy and reliability of media profiling efforts. Our approach introduces a sophisticated combination of a recursive gating mechanism and depthwise separable convolution, which are employed to extract multi-scale local and global association features between factual and biased features. By fusing implicit associations through an attention mechanism, our proposed method effectively mines and models these associations to assist in subsequent classification. This methodology ensures that both granular and comprehensive aspects of the data are captured, thereby enriching the feature set. By integrating these features and enhancing context awareness through the application of multi-level and multi-dimensional attention mechanisms, our framework achieves a more nuanced and accurate representation of the media landscape. Unlike previous work that often assessed tasks independently, our model overcomes these limitations by adopting a complex neural network structure to extract hidden information. Furthermore, it innovatively uses correlation features to assist in profiling media sources, reducing the difficulty and cost of implementation compared to graph-based methods. Extensive experimental verification demonstrates that combining bias and factuality features significantly improves the performance of the media profiling model. This research first applied experiments on bias and factuality assessment tasks on multiple datasets, and compared our proposed framework with various state-of-the-art methods, indicating its superiority. Subsequently, further experimental analysis is conducted and UMAP is used for feature dimensionality reduction visualization to visually demonstrate the significant improvement in the classification performance of

the profiling task after introducing our framework. Additionally, ablation experiments conducted on the structure of the proposed framework discuss the importance and necessity of each module.

In future work, we will explore multi-modal methods to incorporate diverse data sources and enhance the interpretability of our representations. We hope to further refine the performance of news media profiling, ultimately contributing to the creation of a more credible, fair, and sustainable news dissemination environment.

Author Contributions: The authors confirm contributions to the paper as follows: conceptualization: Qi Wang; methodology: Qi Wang and Chenxin Li; software: Chichen Lin; validation: Qi Wang and Chenxin Li; formal analysis: Weijian Fan and Shuang Feng; investigation: Chichen Lin; resources: Yuanzhong Wang; data curation: Chenxin Li; writing—original draft preparation: Qi Wang; writing—review and editing: Qi Wang and Chenxin Li; visualization: Weijian Fan; supervision: Shuang Feng; project administration: Qi Wang; funding acquisition: Yuanzhong Wang. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the corresponding author, Chenxin Li, upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

[1] W. Fan, Y. Wang, and H. Hu, "Mimicking human verification behavior for news media credibility evaluation," *Appl. Sci.*, vol. 13, no. 7, Aug. 2023, Art. no. 9553. doi: 10.3390/app13179553.

[2] G. Song and Y. Wang, "Mainstream value information push strategy on Chinese aggregation news platform: Evolution, modelling and analysis," *Sustainability*, vol. 13, no. 19, Oct. 2021, Art. no. 11121. doi: 10.3390/su131911121.

[3] M. Spliethöver, K. Maximilian, and H. Wachsmuth, "No word embedding model is perfect: Evaluating the representation accuracy for social bias in the media," in *Findings of the Assoc. for Computat. Linguist., EMNLP 2022*, Abu Dhabi, United Arab Emirates, Dec. 2022, pp. 2081–2093.

[4] M. Abulaish, A. Saraswat, and M. Fazil, "A multi-task learning framework using graph attention network for user stance and rumor veracity prediction," in *Proc. IEEE/ACM Int. Conf. Adv. in Soc. Netw. Anal. Mining*, Kusadasi, Turkiye, 2024, pp. 149–153.

[5] N. Luo, D. Xie, Y. Mo, F. Li, C. Teng and D. Ji, "Joint rumour and stance identification based on semantic and structural information in social networks," *Appl. Sci.*, vol. 54, pp. 264–282, Dec. 2024. doi: 10.1007/s10489-023-05170-7.

[6] G. Ma, C. Hu, Li Ge, and H. Zhang, "DSMM: A dual stance-aware multi-task model for rumour veracity on social networks," *Inf. Process. Manag.*, vol. 61, no. 1, Jan. 2024, Art. no. 103528. doi: 10.1016/j.ipm.2023.103528.

[7]   N. Bai, F. Meng, X. Rui, and Z. Wang, "A multi-task attention tree neural net for stance clas-sification and rumor veracity detection," *Appl. Sci.*, vol. 53, pp. 10715–10725, Aug. 2023. doi: 10.1007/s10489-022-03833-5.

[8]   L. Chen et al., "Joint stance and rumor detection in hierarchical heterogeneous graph," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 6, pp. 2530–2542, Jun. 2022. doi: 10.1109/TNNLS.2021.3114027.

[9]   X. Han, H. Zhen, M. Lu, D. Li, and J. Qiu, "Rumor verification on social media with stance-aware recursive tree," in *Proc. Knowl. Sci., Eng. Manag.*, Cham, Germany, Aug. 07, 2021, pp. 149–161.

[10] R. Yang, J. Ma, H. Lin, and W. Gao, "A weakly supervised propagation model for rumor verification and stance detection with multiple instance learning," in *Proc. 45th Int. ACM SIGIR Conf. Res. Dev. in Inf. Retrieval*, Madrid, Spain, Jul. 2022, pp. 1761–1772.

[11] R. Baly, G. Karadzhov, D. Alexandrov, J. Glass, and P. Nakov, "Predicting factuality of reporting and bias of news media sources," in *Proc. Empirical Methods Nat. Lang. Process.*, Brussels, Belgium, Oct.–Nov. 2018, pp. 3528–3539.

[12] R. Baly et al., "What was written vs. who read it: News media profiling using text analysis and social media context," in *Proc. 58th Annu. Meet. Assoc. Computat. Linguist.*, Jul. 2020, pp. 3364–3374.

[13] P. Carragher, E. M. Williams, and K. M. Carley, "Detection and discovery of misinformation sources using attributed webgraphs," 2024, *arXiv:2401.02379*.

[14] Y. Lei, R. Huang, L. Wang, and N. Beauchamp, "Sentence-level media bias analysis informed by discourse structures," in *Proc. Empirical Methods in Nat. Lang. Process.*, Abu Dhabi, United Arab Emirates, Dec. 2022, pp. 10040–10050.

[15] P. Panayotov, U. Shukla, H. T. Sencar, M. Nabeel, and P. Nakov, "GREENER: Graph neural networks for news media profiling," in *Proc. Empirical Methods in Nat. Lang. Process.*, Abu Dhabi, United Arab Emirates, Dec. 2022, pp. 7470–7480.

[16] M. Gruppi, B. D. Horne, and S. Adalı, "Tell me who your friends are: Using content sharing behavior for news source veracity detection," in *Proceed. Preprint (Preprint'21)*. ACM, New York, NY, USA, 2021. doi: 10.48550/arXiv.2101.10973.

[17] M. Nikhil, P. M. Leonor, and G. Dan, "Tackling fake news detection by continually improving social context representations using graph neural networks," in *Proc. Annu. Meet. Assoc. for Computat. Linguist.*, Dublin, Ireland, May 2022, pp. 1363–1380.

[18] B. Gawronski, "Partisan bias in the identification of fake news," *Trends Cogn. Sci.*, vol. 25, pp. 723–724, Jul. 2021. doi: 10.1016/j.tics.2021.05.001.

[19] A. Khandelwal, "Fine-tune longformer for jointly predicting rumor stance and veracity," in *Proc. 3rd ACM India Joint Int. Conf. Data Sci. & Manag. Data (8th ACM IKDD CODS & 26th COMAD)*, Bangalore, India, Jan. 2021, pp. 10–19.

[20] P. Wei, N. Xu, and W. Mao, "Modeling conversation structure and temporal dynamics for jointly predicting rumor stance and veracity," in *Proc. Empirical Methods Nat. Lang. Process. 9th Int. Joint Conf. Nat. Lang. Process. (EMNLP-IJCNLP)*, Hong Kong, China, Nov. 2019, pp. 4787–4798.

[21] Z. Chen, S. C. Hui, L. Liao, and H. Huang, "SSRI-Net: Subthreads stance-rumor interac-tion network for rumor verification," *Neurocomputing*, vol. 583, May 2021, Art. no. 127549. doi: 10.1016/j.neucom.2024.127549.

[22] J. D. Krieger, T. Spinde, T. Ruas, J. Kulshrestha, and B. B. Gipp, "A domain-adaptive pre-training approach for language bias detection in new," in *Proc. 22nd ACM/IEEE Joint Conf. Digit. Libraries*, Cologne, Germany, Jun. 2022, pp. 1–7.

[23] A. M. Luvembe, W. Li, S. Li, F. Liu, and X. Wu, "CAF-ODNN: Complementary attention fusion with optimized deep neural network for multimodal fake news detection," *Inf. Process. Manag.*, vol. 61, no. 3, May 2024, Art. no. 103653. doi: 10.1016/j.ipm.2024.103653.

[24] Y. Zhou, Y. Yang, Q. Ying, Z. Qian, and X. Zhang, "Multi-modal fake news detection on social media via multi-grained information fusion," in *Proc. 2023 ACM Int. Conf. Multimedia Retrieval*, Thessaloniki, Greece, Jun. 2023, pp. 343–352.

[25] Y. Liu, X. F. Zhang, D. Wegsman, N. Beauchamp, and L. Wang, "POLITICS: Pretraining with same-story article comparison for ideology prediction and stance detection," in *Findings of the Association for Computat. Linguist.: NAACL 2022*, Seattle, USA, Jul. 2022, pp. 1354–1374.

[26] E. Diego, A. J. Reddy, P. Chawla, and J. Lehmann, "Belittling the source: Trustworthiness indicators to obfuscate fake news on the web," in *Proc. Fact Extraction and Verification (FEVER)*, Brussels, Belgium, Nov. 2018, pp. 50–55.

[27] B. Sergio, S. Dairazalia, V. Esaú, and M. Petr, "Reliability estimation of news media sources: Birds of a feather flock together," in *Proc. North Am. Chapter Assoc. Computat. Linguist.: Hum. Lang. Technol.*, Mexico City, Mexico, Jun. 2024, pp. 6900–6918.

[28] Hounsel *et al.*, "Identifying disinformation websites using infrastructure features," 2020, *arxiv:2003.07684*.

[29] Y. Kai-Cheng and M. Filippo, "Large language models can rate news outlet credibility," 2023, *arXiv:2304.00228*.

[30] M. Luo, J. T. Hancock, and D. M. Markowitz, "Credibility perceptions and detection accuracy of fake news headlines on social media: Effects of truth-bias and endorsement cues," *Commun. Res.*, vol. 49, no. 2, pp. 171–195, 2022. doi: 10.1177/0093650220921321.

[31] M. Babaei *et al.*, "Analyzing biases in perception of truth in news stories and their implications for fact checking," *IEEE Trans. Comput. Soc. Syst.*, vol. 9, no. 3, pp. 839–850, Jun. 2022. doi: 10.1109/TCSS.2021.3096038.

[32] Y. Dai, F. Gieseke, S. Oehmcke, Y. Wu, and K. Barnard, "Attentional feature fusion," in *2021 IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Waikoloa, HI, USA, 2021, pp. 3559–3568. doi: 10.1109/WACV48630.2021.00360.

[33] Y. Rao *et al.*, "HorNet: Efficient high-order spatial interactions with recursive gated convolutions," in *Proc. Neural Inf. Process. Syst. (NIPS '22)*, New Orleans, LA, USA, Nov. 2024, p. 752.

[34] J. Deng, L. Cheng, and Z. Wang, "Attention-based BiLSTM fused CNN with gating mechanism model for Chinese long text classification," *Comput. Speech & Lang.*, vol. 68, Jul. 2021, Art. no. 101182. doi: 10.1016/j.csl.2020.101182.

[35] C. Liu and X. Xu, "AMFF: A new attention-based multi-feature fusion method for intention recognition," *Knowl.-Based Syst.*, vol. 233, Dec. 2021, Art. no. 107525. doi: 10.1016/j.knosys.2021.107525.

[36] X. Ding and Y. Mei, "Research on short text classification method based on semantic fusion and BiLSTM-CNN," in *Proc. Inf. Sci., Electr., and Automat. Eng. (ISEAE 2022)*, Aug. 2022, pp. 538–543.

[37] M. Lin, Q. Chen, and S. Yan, "Network in network," in *Proc. Int. Conf. Learn. Represent.*, Dec. 2013.

[38] B. D. Horne, W. Dron, S. Khedr, and S. Adalı, "Assessing the news landscape: A multi-module toolkit for evaluating the credibility of news," in *Proc. Web Conf. 2018 (WWW'18)*, Lyon, France, Apr. 2018, pp. 235–238.

[39] B. D. Horne, J. Nørregaard, and S. Adali, "Robust fake news detection over time and attack," *ACM Trans. Intell. Syst. Technol.*, vol. 11, no. 1, pp. 1–23, Dec. 2019. doi: 10.1145/3363818.

[40] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform manifold approximation and projection for dimension reduction," 2018, *arXiv:1802.03426*.