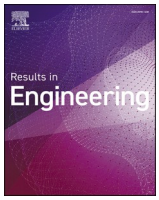
Results in Engineering 26 (2025) 104752

Contents lists available at ScienceDirect

Results in Engineering

journal homepage: www.sciencedirect.com/journal/results-in-engineering



Review article

A systematic review of multimodal fake news detection on social media using deep learning models

Maged Nasser [a,*], Noreen Izza Arshad [a], Abdulalem Ali [b], Hitham Alhussian [a], Faisal Saeed [c], Aminu Da'u [d], Ibtehal Nafea [e]

[a] Department of Computer & Information Sciences, Universiti Teknologi Petronas, 32610, Bandar Seri Iskandar, Perak, Malaysia
[b] Institute of Computer Science and Digital Innovation, UCSI University, Federal Territory of Kuala Lumpur, Malaysia
[c] DAAI Research Group, Department of Computing and Data Science, School of Computing and Digital Technology, Birmingham City University, Birmingham B4 7XG, UK
[d] Department of Computer Science, Hassan Usman Katsina Polytechnic, Katsina State, Nigeria
[e] College of Computer Science and Engineering, Taibah University, Medina 41477, Saudi Arabia

ARTICLE INFO

Keywords:
Multimodal fake news detection
Deep learning models
Transformers
Recurrent neural network (RNN)
Convolutional neural networks (CNNs)
Autoencoder (AE)

ABSTRACT

The volume of data circulating from online sources is growing rapidly and comprises both reliable and unreliable information published through many different sources. Researchers are making plausible efforts to develop reliable methods for detecting and eliminating fake web news. Deep learning (DL) methods play a vital role in addressing various fake news detection problems and are found to perform better compared to conventional approaches, making them state-of-the-art in this field. This paper provides a comprehensive review and analysis of existent DL-based models for multimodal fake news detection, focusing on diverse aspects, including user profiles, news content, images, videos, and audio data. This study considered the latest articles within the last seven years, starting from 2018 to 2025, and about 963 quality articles were obtained from the journals and conferences selected for this study. Subsequently, 121 studies were chosen for our SLR after careful screening of the abstract and the full-text eligibility analysis. The findings showed that the Transformer models and Recurrent Neural Networks (RNNs) are the most popular deep learning techniques for detecting multimodal fake news, followed by the Convolutional Neural Networks (CNNs) techniques. The Twitter and Weibo datasets are the two most frequently used standard datasets, and the most frequently used metrics to evaluate the performance of these models are the accuracy, precision, recall, and F-scores. In conclusion, the limitations of the current methods were summarized and some exciting possibilities for future research were highlighted, including designing robust multilingual fake news detection systems, hybridization of deep learning models to enhance detection accuracy, integration of explainable AI (XAI), and facilitating real-time fake news detection models.

1. Introduction

In recent times, the internet has influenced how individuals engage and communicate due to its ease of accessibility, low cost, and rapid transmission of data. As a result, many individuals now prefer using social media and online portals to read and search for news of interest rather than traditional newspapers. Although social media has developed into a potent informational tool, it influences society by having a significant impact on major events [1]. The amount of information shared through web content is rapidly increasing, including misinformation that is published by different sources. Further, many of the recent news articles on social media are found to be deliberately misleading [2,3]. Such malevolent news pieces are referred to by a more general phrase, "fake news." False or biased news reports produced for personal benefit mislead readers with deceptive content, which has a significant impact on public opinion and societal stability [4,5]. For example, the widespread fear and anxiety surrounding COVID-19 led to the emergence of psychological and physical illnesses as well as lowered immunity in the general public [6,7]. The 2016 US presidential election is another event that may have been impacted by fake news on social media. Consequently, this expression has spread across society and is now used often. Social media platforms make it simple to create and

* Corresponding author.
E-mail addresses: maged.nasser@utp.edu.my (M. Nasser), noreenizza@utp.edu.my (N.I. Arshad), Almaldolah2012@gmail.com (A. Ali), seddig.alhussian@utp.edu.my (H. Alhussian), fiasal.saeed@bcu.ac.uk (F. Saeed), inafea@taibahu.edu.sa (I. Nafea).

https://doi.org/10.1016/j.rineng.2025.104752
Received 14 December 2024; Received in revised form 13 March 2025; Accepted 28 March 2025
Available online 9 April 2025

disseminate news content, which generates a lot of data to analyze. It is difficult to detect false news technologically for several reasons. Fake news identification is made more difficult by the abundance of content available online that covers a wide range of subjects. Many researchers are now focusing on automated false news identification as a result of this. To this effect, more researchers have put forward a number of machine learning and deep learning models in order to achieve this.

Computer vision and natural language processing (NLP) are two areas where deep learning approaches have completely changed the landscape [8]. The ability to identify toxic language and misinformation in multimedia data is one area where this breakthrough has been very beneficial. Through the integration of both textual and visual data, scientists have created robust algorithms that can distinguish harmful and misleading language on online platforms. Due to its capacity to discover patterns from data, deep learning approaches have attracted a lot of interest lately [9]. In the area of natural language processing, their significance has grown. In a variety of applications, including sentiment analysis, language comprehension, and computer vision, these methods have shown considerable potential. Using deep learning to identify multimodal false news has gained popularity in the last several years. This is a result of growing worries around the spreading of deceptive data on social networks. This method scrutinizes many data sources and unearths misleading information by combining audio analysis, computer vision, and natural language processing. To monitor development and spot obstacles, studies that examine various methods and frameworks used in this field are crucial. Through a comprehensive analysis of several deep learning strategies pertaining to multimodal fake news detection, scholars may discern the advantages and drawbacks of current approaches, acquire further knowledge, and suggest more inventive and superior solutions [10].

Fake news detection on social media presents a number of unique and challenging research issues. While misinformation is not a new problem groups or countries have historically utilized media for propaganda or The emergence of web-generated news on social media amplifies its potency, undermining traditional journalistic standards [11]. This issue has numerous aspects that make automated detection particularly tough. The primary feature of false news is its purposeful design to deceive readers, making detection based only on content nontrivial. The substance of misinformation is varied in themes, methods, and media channels, aiming to distort the facts using various language approaches while concurrently ridiculing authentic news.

To understand the current state of the research on fake news detection approaches and several issues surrounding the field, this study begins by presenting a few definitions for better clarification of the deceptive content, which is necessary to better guide the future directions. In order to better understand the contributions and contents of the research articles, the study also systematically generated research papers from reputable journals and conferences. Then, using recently created standard datasets, it critically evaluated the performance of several existing DL based models for multi-modal fake news detection in social media. Furthermore, we provide a current taxonomy of multimodal false news detection (MFD) features and fusion methods, which is still an open issue that requires further research. Therefore, investigating these issues is essential to have a thorough insight of the existing in the subject and open up new avenues for study. Therefore, the following highlights the main contributions of this study that differentiate it from previous studies:

- We have conducted a comprehensive and current evaluation of multimodal false news detection using deep learning methodologies. We have categorized the recent publications based on several deep learning approaches, including RNN, Recursive Neural Networks (RvNNs), CNNs, Auto Encoder (AE), Generative Adversarial Network (GAN), hybrid models, and the newest state-of-the-art methods.
- We provide a current taxonomy of features for fake news detection thereby offering several feature combinations for multimodal fake news detection methods.
- We have analyzed all the available datasets for multimodal fake news detection and reviewed the performance of the suggested models by researchers on these datasets.
- We outline the evaluative metrics used to evaluate the performance of DL models employed for MFD.
- This study is concluded by presenting a summary of research findings, along with a critical evaluation of future research studies and other unresolved issues or research gaps.

The rest of this paper is structured as follows: Section 2 present the overview of fake news detection. Section 3 introduces the methodology adopted in the systematic literature review; Section 4 presents the results which answer the research questions designed for this study. Section 5 presents the limitation of the study, while Section 6 concludes the study by noting its scientific achievements and outlining future challenges that require resolution.

2. Materials and methods

In this review, the guiding structure for efficient systematic literature review writing (SLR) depends on the famous Precise Surveys and Meta-Examinations (PRISMA) rules [12]. PRISMA offers a consistent and reproducible process for finding, selecting, and evaluating current studies. It also guides the selection, identification, and assessment of the studies. The following subsection provides more information on the review process, shown in Fig. 1.

2.1. Search strategy and data sources

Eight different bibliographic databases were searched to locate more pertinent works. Web of Science, Scopus, Google Scholar, the ACM Digital Library, SpringerLink, ScienceDirect, and the IEEE Xplore Library are just a few of the digital libraries that were searched. The choice of these sources is based on their sufficiency in providing rich information regarding research articles and their popularity to be considered by several researchers as worthy of reliable exploration. Other similar sources were not considered as they mainly index data from the primary sources. The evaluation period considered for this study is from 2018 to 2025 to get the most recent and comprehensive review. As shown in Fig. 2, we additionally employ word synonyms and logical operators

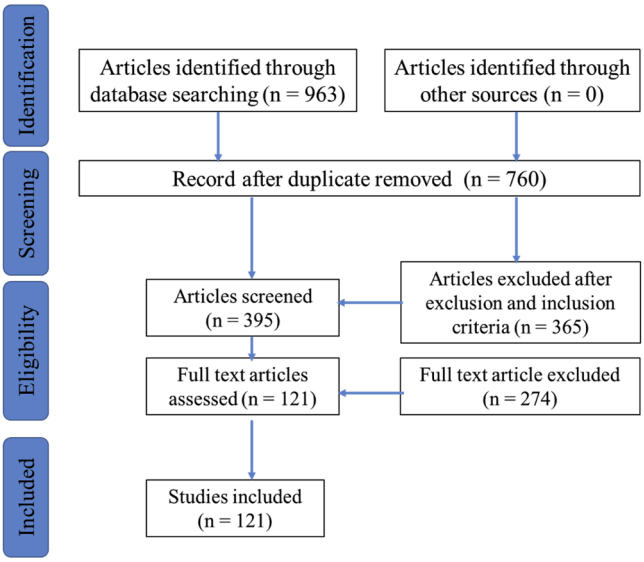


Fig. 1. PRISMA approach for the SLR.

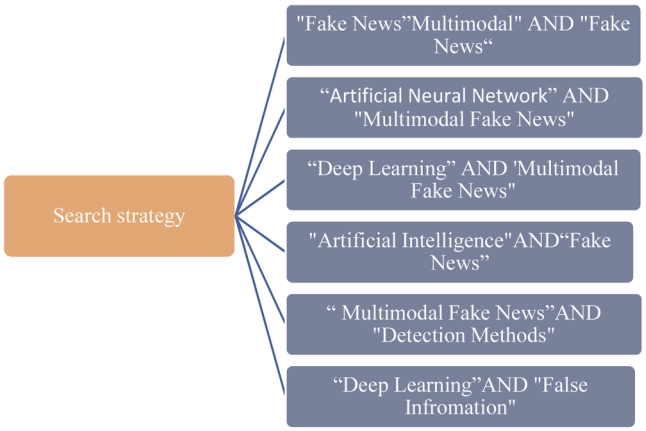


Fig. 2. Search strategy.

("OR," "AND"). The search string consists of "Fake News" or "False Information" or "Multimodal Fake News" "((Artificial Neural Network)" or "((deep learning) AND (Multimodal Fake News))" or "((Artificial Intelligence) AND (Fake News) AND (detection techniques))" or "Fake News Detection")). Using the chosen keywords to conduct the searching of the relevant articles, a substantial amount of published articles was obtained related to FND in different aspects.

2.2. Selection criteria

To fully identify the relevant articles, we perform a manual search along with the search phrase used in the automatic search. There were over 1200 studies obtained in total through the identification stage. After removing duplicate, unsuitable, and irrelevant papers from the screening stage, 963 published papers were selected. After going through the abstract, introduction, and title, 203 published studies were eliminated. The remaining 760 were then subjected to the removal criteria, and 365 articles were eliminated. 395 research articles were then advanced to the next phase. Following the review of the entire papers in the eligibility phase, 274 research publications were dropped, and 121 research articles were considered for this SLR of DL-based models for fake news detection.

The number of publications was limited to 121 articles by taking into account only the articles that explicitly focused on existing DL-methods and those articles that were exclusively from indexed journals and conferences. Table 1 lists the exclusion/inclusion criteria for the SLRs.

2.3. Quality assessment

In SLR research, evaluating the data and the value of the evidence it provides is critical. Since biases resulting from the research processes can have impacts on the results of a poorly conducted study, it is important to carefully analyze the results of the study. Irrelevant papers are required to be explicitly excluded from the systematic review, or at the very least identified as such. Moreover, it is important to select the appropriate standards for evaluating the quality of the findings and any possible biases in the selected study.

Standard quality checklist questions (SQCQ) created by [13] are used to validate the quality of the chosen publications for this study. To do this, we chose the papers that responded "yes" to a minimum of seven questions, as specified in [14]. The quality evaluation will be considered in addition to the data extraction to ensure the research findings have an important effect on the review [15]. Table 2 displays the SQCQ employed in this study.

Table 1
The inclusion and exclusion criteria for the SLRs.

No.	Inclusion	Exclusion
1	Articles that only involve experimental studies	Articles that do not involve experimental studies were not regarded
2	Articles published from 2018 to 2025	Articles published earlier than 2018 were excluded
3	Articles involving fake news detection only	Articles involving detecting methods other than fake news detection.
4	Research on identifying fake news with deep learning	Study identifying additional methods for detecting fake news
5	Only papers written in English-language	Papers written in a different language
6	Only articles from journals and conferences are considered.	Books, theses, and magazines are not considered.

2.4. Data extraction and synthesis

Most importantly, we noted the vital information contained in the selected articles, such as the journal title, the year it was published, the list of authors, and the publisher. The accuracy of the deep learning techniques reported, and evaluation metrics were also added as data for the SLR study. The results of the data extraction process that produced the studied articles can be applied to the research questions specifically designed in the next section to fulfill the primary goal of the SLR. To properly analyze research findings, we used a variety of visualization techniques and tools, including line charts, pie charts, histograms, etc., to visualize and evaluate the data.

2.5. Research questions

Determining the overall goal and expected results of a study requires careful consideration when choosing RQs. As a result, we created the following RQs to fulfill the primary goal of our SLR:

- RQ-1: What are the recent DL methods used for multi-modal fake news detection problems?
- RQ-2: What are the features commonly utilized in the multimodal fake news detection with DL models?
- RQ-3: What are the standard datasets commonly utilized in the literature for evaluation of the DL models for the fake news detection?
- RQ-4: What are the common performance metrics for fake news detection models using DL?
- RQ-5: What are the challenges and prospects of DL-based fake news detection approaches?

3. Multimodal fake news detection

Essentially, fake news detection is the process of identifying and labelling news content as fake or real. In order to combat misrepresentation and promote the spreading of correct information, FND aims to provide systems that can specifically recognize and label fake news documents [16–21]. The majority of works in this topic have examined

Table 2
Standard Quality Checklist Questions.

No.	Standard Questions
1	Is the description coherent and clear?
2	Is the purpose of the study effectively stated?
3	Is the data gathering process well explained?
4	Have the various contexts been sufficiently investigated?
5	Is the study's findings reliable?
6	Are the data, analysis, and conclusion related?
7	Is the experimental process and method understood?
8	Are the study approaches well documented?
9	If important, are they reliable?
10	Can the findings of the research be repeated?

the challenge of binary classification in the literature with regard to identifying false news. Thus, the following general concept of fake news detection is developed:

$$F(a_i) = \begin{cases} 0 \text{ if } x \text{ is a fake new} \\ 1 \text{ otherwise} \end{cases}$$

Using the features extracted from the news article x, the fake news detection model aims at classifying the news article as genuine or false. F represents detection model for detecting the fake news. A feature that was extracted from news item x is represented by a*i*, where $i = 1, 2, 3$.

Consequently, several studies have been introduced on multimodal fake news detection (MFND) approaches. Unlike single-modal detection approach, MFND approach considers multiple features, such as news contents, user profile, audio and visual data to improve detection accuracy. In the studies such as [22–28], authors proposed different techniques, ranging from transformer models, language models, and other aspects such as the writing styles employed in the news, the news's spread pattern, and the legitimacy of the news source used. The success of these techniques increases the interest of researchers on this field and more works were presented.

As illustrated in Fig. 3, the progression of research interests in this area is presented, showing the chronology of the papers that were chosen for this SLR and the number of articles covering multimodal false news detection methods based on deep learning from 2018 to 2025. The illustration in Fig. 3 shows a rising trend of research in this area over the past few years, particularly starting in 2019 with a sharp increase in the number of articles published. The announced wealth of fake news encompassing the Coronavirus pandemic via online entertainment might have added to this expanded interest in counterfeit news ID [29]. Along these lines, the year 2020 and 2021 saw a few papers (12 and 13 papers), while, the 2023 and 2024 recorded the most papers distributed in the review years (21 and 23 articles), respectively. This signifies the relevance of this topic to the research communities as the study progresses over time.

Based on the analyzed articles, it was found that multiple feature combinations were investigated in the context of multi-modal false news identification. Recent studies conducted in MFND combined different types of feature sets, such as the fusion of visual and textual features, video and textual features, video and audio features, and user profile with news content features. Even though these approach demonstrate strong capabilities, very few take into account a broader range of features from news content and social contexts for more accurate detection. The literature indicates that more research is needed for determining the differences in the detection approaches and efficiency of various detection algorithms, feature combinations and performance measures of evaluations.

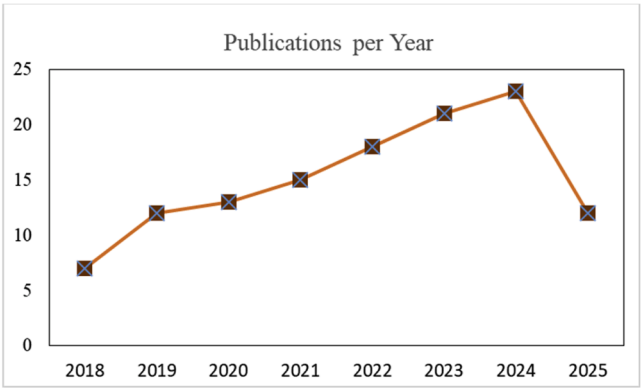


Fig. 3. Number of publications per year.

3.1. Available features for fake news detection

The features derived from the news articles may be applied for identifying fake news, as found in the literature. Depending on how many features are present, the task of detecting fake news is to classify it as multi-modal or single-modal [2]. Fig. 4 shows the taxonomy of different categories of fake news detection methods that focused on feature-based approaches.

In the single-modal approach, the features utilized for FND are categorized into two classes: social context feature and news content features [11]. Fig. 5 depicts the taxonomy of characteristics accessible in a single-modal application for detecting false news. The content-based method aims to identify patterns within the elements of news articles extracted from their main body. These content-based features are categorized into two groups: visual features and textual features [29]. The textual features can be studied from various aspects based on the features such as the style-based, linguistic-syntactic features, and context structure features. The visual-based features include hidden distributions in the image, video and audio news contents. Explanation for single modal features can be found in [2].

3.2. Data modalities for multimodal FND

Multimodal systems are more resilient to sophisticated fake news schemes, such as deepfakes, which often involve multiple types of media. By analyzing all available features, these models can detect subtle signs of manipulation that might be missed when considering a single modality. This section describes different feature combinations, such as the fusion of text and visual features, text and video features, video and audio features, and user profiles with news content features to develop MFND methods. Table 3 shows the description of the features used for multimodal fake news detection models and Fig. 6 shows the distribution of different features used for multimodal fake news detection based on the selected research studies.

3.2.1. Textual and visual features

The most often utilized combination of modalities for detecting false news using deep learning algorithms is the mix of textual and visual elements. NLP methods are used by numerous multimodal fake news detection algorithms to process and analyze massive amounts of text to extract the text characteristics and identify false news [30]. Word embedding, feature extraction approaches, and pre-processing are some of the activities involved in this procedure. Typically, the first stage is data pre-processing, which includes managing missing words, binarizing attributes, expressing ambiguous attributes, and handling complex

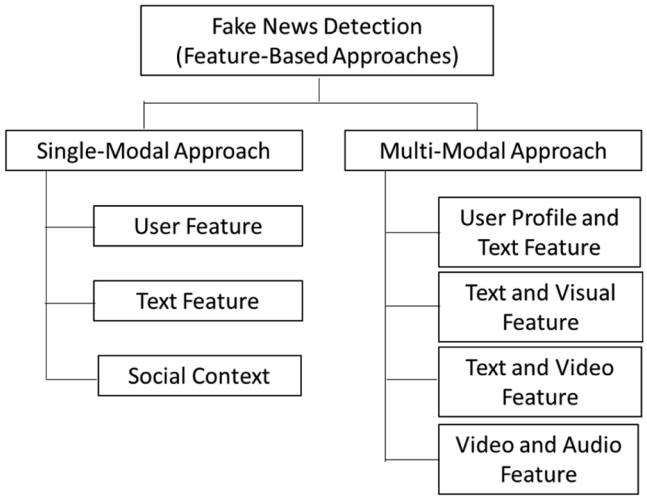


Fig. 4. different features utilized for FND.

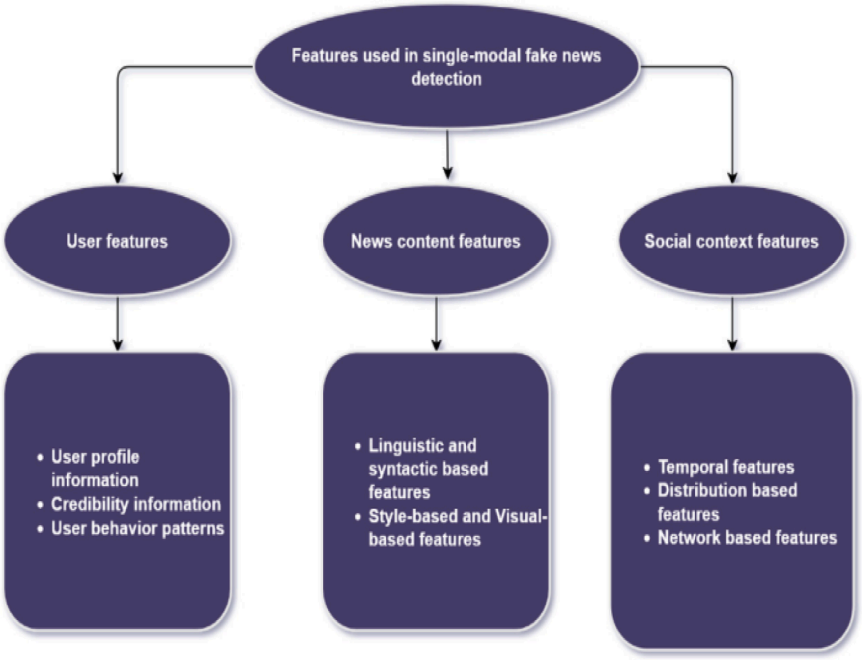


Fig. 5. Features used in single-modal FND [2].

structures including attributes. Different visualization approaches are helpful in the pre-processing stage of the data. When resolving noisy data, data pre-processing saves space and lowers computer overhead. Secondly, word vectorization refers to converting text or words into a set of vectors. Additionally, various ML methods frequently employ TF-IDF along with a bag-of-words model for identifying false information [31, 32]. False news detection algorithms have recently begun to leverage pre-trained embedding models such as word2vec and GloVe because they can train larger datasets [33].

Words with multiple n-gram orders may be generated by N-gram [34]. In other words, each informational N-gram vector is clustered to get a single feature vector. Linguistic feature extraction [9] is used to examine the performance of fake news. It comprises several feature classes, such as quantity features, user credibility, stylistic characteristics, psycho-linguistic features, and readability index. For the activities that follow, word embedding is used to create word vectors [35]. Nevertheless, it might be difficult to start from scratch when making word vectors out of several words on a large dataset. BERT has also been widely used for a wide range of additional NLP tasks, including text classification, translation, and summarization[36]. Due to its context awareness, it does very well on tasks requiring deep language comprehension [25,37]. Then, the image features are mostly extracted by using the VGG-19 technique [37–39]. It is also important to normalize the size of images and divide them for training and testing. Textual and visual data are pre-processed individually and then merged to complete each instance in terms of its three parameters: title, text, and vision [40].

3.2.2. *Text and video*

In addition to textual data, multimodal fake news detection leverages various types of data, such as videos, to improve the accuracy of identifying false information. Videos can be analyzed for inconsistencies between the visual content and the accompanying text or audio. For example, mismatched subtitles or altered video frames can be red flags [38,41,42]. On this note, [41] proposed a Multimodal Neural Network (MNN) which focuses on identifying fake news by analyzing the alignment between different types of information, such as text and images. The proposed MCNN model incorporates five subnetworks including text highlight extraction, visual semantic element extraction, visual altering discovery, closeness estimation, and multimodal combination. The message and visual semantic component extraction modules map the message and visual elements to a typical space for cross-modular portrayal. The visual altering highlight extraction module recognizes physical and altering highlights in pictures. In light of this methodology, the likeness estimation module surveys the arrangement of text and pictures to distinguish irregularities. To assess the presentation of the proposed model. The four generally utilized counterfeit news identification datasets were utilized. The outcome shows the superior exactness contrasted with existing strategies, exhibiting its viability in recognizing counterfeit news by guaranteeing the consistency of multimodal information.

In addition, [38] introduce a comprehensive approach to identifying fake news based on the SpotFake model for multimodal fake news detection. The Spot Fake model integrates text and image data to detect fake news without relying on additional subtasks like event discrimination. The proposed model was tried on two freely accessible Twitter and Weibo datasets, and the outcome shows that SpotFake beat existing cutting-edge techniques as far as precision on both datasets [42] proposes the CB-Fake model, designed to detect fake news by integrating both textual and visual features. The CB-Fake model utilizes Capsule Neural Networks to extract key visual features from images and BERT to capture semantic relationships within text. By combining these two approaches, the model creates a multimodal feature vector that strengthens the information content, improving the overall detection accuracy. This method aims to overcome the shortcomings of earlier techniques by maintaining semantic word correlations and capturing intricate visual details. By evaluating two publicly accessible datasets namely Politifact and Gossipcop, the performance demonstrates the effectiveness of combining advanced neural network models to enhance robustness of FND systems.

3.2.3. *Audio and video*

Detecting fake news using both text and video involves analyzing the consistency and coherence between the textual content and the visual elements of a video [43–49]. On this note, [43] introduced an advanced approach for detecting fake news videos on YouTube by leveraging domain knowledge and multimodal data fusion. Their proposed model combines both machine learning and DL methods to process and integrate data from various modalities. The detection process involves a

Table 3
Description of the features used for MFND.

Data used	Datasets	Feature representation techniques	Fusion method	Reference
Textual and visual features	Weibo Twitter FakeNewsNet Fakeddit	BERT,	Similarity Measurement Module, Multimodal Fusion Module	[41]
	FakeNewsNet	BERT, Pre-train model on ImageNet	Transfer Learning and Multimodal Fusion	[38]
	Politifact Gossipcop	BERT, CapsNet	End-to-End Framework	[42]
	Weibo Twitter	Sentiment analysis, Segmentation method	Cultural algorithm	[55]
	FakeNewsNet	BERT and VGG-16		[40]
	Social-Media	BERT and VGG-16		[56]
	PolitiFact, GossipCop, and Twitter datasets	–	Three-level feature matching distance (TLFND)	[57]
	Twitter and Weibo	BiLSTM, -CNN–RNN	Multimodal Bilinear Pooling (MBP) for feature fusion	[58]
	Twitter Weibo	BERT and VGG-19	Co-Attention Mechanism	[59]
	FakeNewsNet	BERT, ResNet-50, and Bi-LSTM	–	[60]
	FakeNewsNet	multi-level encoding network and VGG-16	Relationship-aware attention network	[61]
	Weibo	BERT and ResNet-50	–	[62]
	Twitter Weibo	BERT and VGG-19	–	[38]
	WEIBO and PHEME	BERT, and ResNet-50	Multi-Modal Cross-Attention Network	[63]
	Twitter Weibo	multi-level encoding network and VGG-19	Relationship-aware Attention Network	[64]
	Fakeddit, Politifact, Gossipcop	MacBERT and MDVT-MS	DeBertNeXT	[65]
	Weibo Twitter FakeNewsNet	multi-level encoding network and VGG19	relationship-aware attention	[66]
	Twitter Weibo	BERT and ResNet-50	Dual Attention Fusion Networks	[67]
	FakeNewsNet	BERT and ViT	Attention Fusion Networks	[68]
	Weibo Twitter	Variational AE and CNN	Bimodal variational autoencoder	[69]
	Fakeddit Flickr	BERT and Xception	Dense fusion network	[70]
Text and Video	YouTube-8M	BERT, CNN, and VGGish	Attention mechanism	[43]
	Misleading Youtube Video Corpus (MYVC)	BERT, CNN, and VGGish	Attention mechanism	[43]
	FakeClips	BERT and ResNet-50	Attention mechanism	[45]
	VAVD and FVC	NLP and CNN	–	[46]
	YouTube-8M	LDA, CNN, VGGish	–	[71]
	YouTube-8M	Biterm Topic Model (BTM), CNN, VGGish	–	[47]
Audio and Video	DeepFake-TIMIT, DFDC (DeepFake Detection Challenge)	CNN, VGGish	Siamese network architecture	[48]
	FakeAVCeleb	Meso-4, MesoInception-4, Xception, EfficientNet-B0, and VGG16.	–	[49]
	FakeAVCeleb	Xception, InceptionResNetV2	Gated Recurrent Unit (GRU) based attention mechanism	[72]
	FakeAVCeleb	CNN, Wav2Lip	Lip-reading-based multimodal approach	[73]
	FDC (DeepFake Detection Challenge) DeepFake-TIMIT	CNN	Modality Dissonance Score (MDS)	[74]
Text, video and Visual	DFDC (DeepFake Detection Challenge) VoxCeleb In-the-Wild (ITW) Audio Deepfakes	CNN, Mel-spectrograms	Syn-Stream	[75]
	FaceForensics++ FDC (DeepFake Detection Challenge) FakeAVCeleb	CNN, EfficientNet-B0, Mel-spectrograms	Attentio mechanism	[76]
	DefakeAVMiT FakeAVCeleb DFDC (DeepFake Detection Challenge)	Temporal-Spatial Encoder	Multi-Modal Joint-Decoder	[77]
	FakeAVCeleb, World Leaders, Presidential Deepfake Detection	CNN	IntrAmodality Mixer and Layer (IAML) IntErModality Mixer Layers (IEML)	[78]
	Fake AVCeleb Celeb- DF ASVSpoof-2019 LA World Leaders Presidential Deepfakes	Swin Transformer, dense network layers	–	[79]
	FDC (DeepFake Detection Challenge) Celeb-DF, FakeAVCeleb	CNN, EfficientNet-B0, Mel-spectrograms	Contrastive Learning	[80]
	FakeSV	BERT, VGG 19	Attention mechanism	[44]
User Profile and News Contents	LIAR BuzzFeed News PolitiFact	Transformer model	Transformer model	[53]
	BuzzFeed PolitiFact	CNN, a coupled matrix-tensor factorization method	A deep neural network	[81]
	LIAR BuzzFeed News	BERT, CNN	Fusion layer	[35]

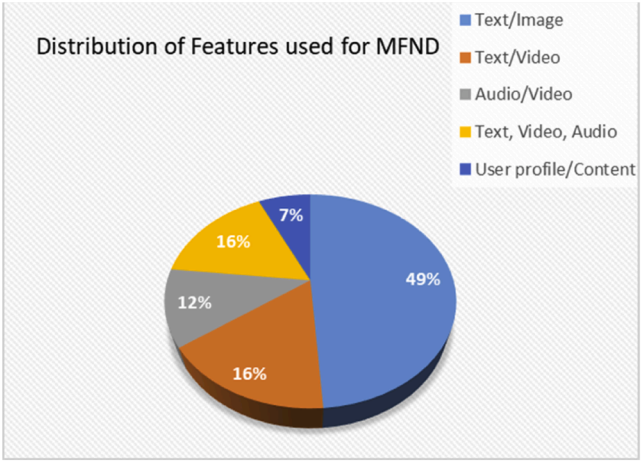


Fig. 6. Distribution of features used for MFND.

two-phase strategy: First, it analyzes textual data using classifiers like BERT, and then it incorporates visual and audio data through a modified Convolutional Neural Network (CNN). When tested on YouTube videos, the framework outperformed baseline models, showing a marked improvement in detection accuracy by efficiently utilizing domain knowledge and multimodal data fusion. The author [45] introduces a deep multitask model that combines fake news detection with emotion recognition as an auxiliary task. The model leverages a newly developed dataset named FakeClips, which includes video-based multimodal fake news data. The model uses supervised contrastive learning for enhancing accuracy. By investigating both emotion and fake news together, the framework aims to capture the emotional appeal often used in fake news to attract viewers. The results show significant improvements in fake news detection in terms of accuracy and F-score. Similarly, [46] address the challenge of identifying misleading metadata in YouTube videos, which can include deceptive titles, descriptions, tags, and thumbnails. The authors developed UCNet, a deep neural network designed to detect fake videos by analyzing metadata and content. UCNet was tested on two datasets, namely VAVD and FVC. The model achieved a macro-averaged F-score of 0.82, significantly outperforming the baseline model.

3.2.4. User profile and news contents

Fake news may be discovered by examining social network information, especially the user-based aspects. User-based characteristics were retrieved from user profiles [50–52]. For example, the number of followers, number of friends, and registration ages are important factors in establishing the authenticity of a user post [37]. Based on this, [53] propose an innovative framework for fake news detection by utilizing both the content of news articles and the social contexts in which they are distributed. In other words, the approach leverages information from the news content and its surrounding social environment to identify misinformation. It is built on a Transformer model, comprising of two modules: an encoder that uses data from false news sources to learn valuable representations and a decoder that uses historical observations to forecast future behavior. To improve our ability to categorize news, it integrates several characteristics from social situations and news content into our model. Furthermore, a successful labeling strategy is used to tackle the issue of label scarcity. Results from experiments using real-world data demonstrate that the suggested approach can identify fake news more quickly and accurately than baselines, even after it has started to spread. Similarly, [54] presents a similar approach to detecting fake news by leveraging both news content and social contexts. This model takes into account the presence of echo chambers, where like-minded users reinforce each other's viewpoints, which plays a key role in detecting fake news. It utilizes a coupled matrix-tensor factorization technique to combine news content with social context, generating a latent representation that improves the detection process. The deep neural network is structured with multiple filters across dense layers and includes dropout for regularization. To assess the presentation, the model was approved on genuine world datasets like BuzzFeed and PolitiFact, accomplishing an approval precision of 92.30 %, beating existing cutting-edge models. Besides, [35] present a BERT-based profound learning framework by incorporating a few equal blocks of the single-layer profound Convolutional Brain Organization (CNN) with changing part sizes and channels with the BERT.

4. The recent DL models used for MFND tasks

In addition to the publication trends, as seen in Fig. 7, it is clear from Fig. 8 that there has been a considerable proliferation in awareness in methods for FND, going from 14 in 2018 to 20 in 2019, with 28 various approaches being used for fake news detection publications in 2021. With 32 distinct ways reported in 2022, 2023 records the most studies, with 46 different strategies in the period from January to August. Additionally, it is clear from Fig. 8 that the deep learning techniques employed for fake news identification have seen varying levels of popularity over time.

Though deep learning models are applicable in other fields [82] and [83], the findings of this review demonstrate that Transformer models, RNN models, and CNN models are the most extensively used DL architectures for the detection of fake news. Although ensemble and GRU techniques have the potential to see more publications because they are more recent in this multimodal fake news detection, their significance remains to be fully appreciated. This section provides a brief description of different DL algorithms for MFND as well as the studies that used it. Table 4 summarizes all the deep learning algorithms utilised for multi-modal fake news detection studies for the specified period.

4.1. Convolutional neural network (CNN)

FND models have recently been introduced to detect fake news across several modalities. These models take advantage of CNNs' strengths in processing visual data and combine them with sophisticated strategies for handling textual information [8,84–86]. For instance, [8] investigates a CNN architecture for multi-modal false news detection that incorporates text and visual data. The outcome shows that using text and picture data together can enhance the efficiency of FND systems. This method performed well for some kinds of false news, which emphasizes how important it is to include a variety of data types in detection algorithms. Similarly, [87] suggests a technique for detecting fake information that uses SE and Text-CNN modules. The study utilized a transformer for enhancing the feature quality and used a convolutional model for processing the text and picture characteristics using BERT technique. A capsule neural network (CapsNet) model was proposed by [42] to effectively identify false news by extracting the most informative visual elements from a picture. By combining these data, a richer data representation is produced that aids in identifying whether the news is authentic or fake. Reference [84] investigated the most advanced techniques for multi-modal online information reliability analysis using DL methods, such as CNNs and RNNs. Without any pre-processing, the models presented demonstrate the rapid improvement in classification tasks. The authors create visual and textual modules to assess their efficiencies across multi-modal datasets, using layers of convolution to leverage latent characteristics found in text and pictures based on CNN models for the identification of false news.

4.2. Recurrent neural networks (RNN)

RNNs are efficient architectures that enable modelling of successive data through the use of network loops. According to recent research, RNNs have demonstrated remarkable success in several tasks. For

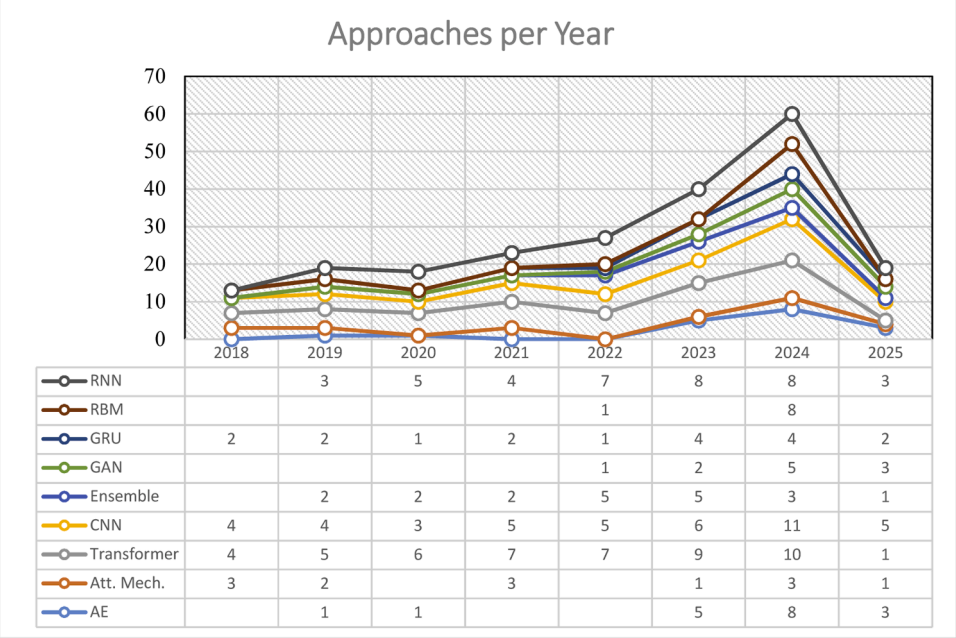


Fig. 7. Number of approaches used by year.

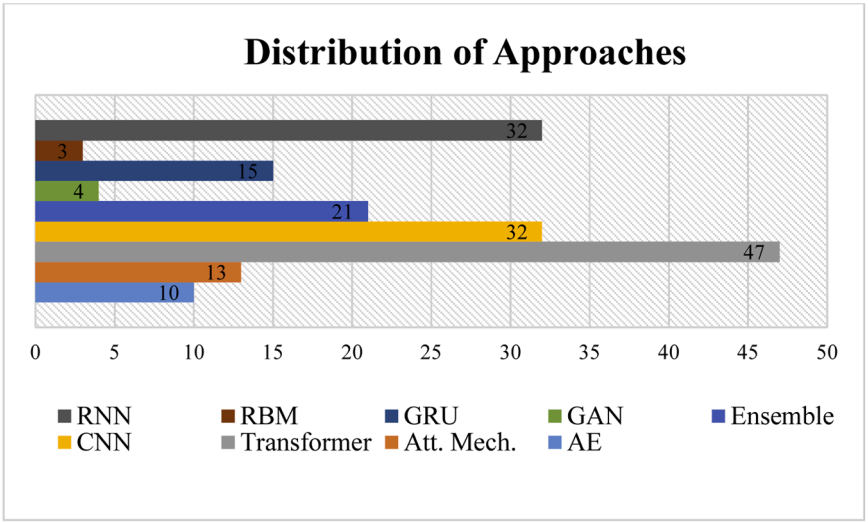


Fig. 8. Comparative distribution of approaches.

example, the capsule neural network (CapsNet) model created by [42] extracts the most informative visual elements from an image. By combining these characteristics, a richer data representation is produced that aids in identifying if the news is authentic or fraudulent. Politifact and Gossipcop, two freely accessible datasets, were utilized to analyze the model's exhibition against different baselines. When contrasted with other state-of-the art calculations, the recommended model gets a lot higher grouping exactness for the Politifact and Gossipcop datasets. Along these lines, [88] presented the idea of a shared consideration brain organization (MANN), which can grasp the associations between different modalities. Reference [41] presented a Multimodal Consistency Brain Organization (MCNN) intended to keep up with the consistency of multimodal information while catching the extensive elements of online entertainment data. This approach comprises several sub-networks which include extraction part, the visual semantic component, the visual altering highlight component and the similitude estimation component. Furthermore, [60] made a multimodal model for identifying misleading news utilizing a more profound learning engineering. This model included text and pictures gotten from online entertainment progressively. With the utilization of glove inserting and Word2vector, the text qualities were recovered. Reference [89] presented another method that mixes consideration processes for design total and logical combination across areas with LSTM networks for literary information translation.

4.3. *Recursive neural network (RvNN)*

To provide an effective performance across variable-size input structures, recursive neural network is a deep neural network that traverse a given structure in topological order while using the same set of weights each time. As a result [90] introduced a deep neural network model that incorporates implicit aspects of both text and visuals to identify multimodal false news. To concentrate on the most relevant portions of the text and visuals, the authors provide an RvNN-based

Table 4
Comparison of deep learning based methods for MFND.

Models	Description	Advantages	Disadvantages	References
Attention Mechanism	Enhance artificial intelligence models by selectively concentrating on essential input components	Attention mechanisms enables focusing on the most informative data, to enhance their understanding context	Attention mechanisms can be computationally expensive, especially for very large datasets or long sequences, as they require calculating attention scores for each pair of input elements	[58,106,85,59, 110,63,95]
AE	An autoencoder is used for unsupervised learning. Its main purpose is learning a lower-dimensional representation of higher-dimensional data. Essentially, it compresses input data into a compact form while retaining essential features.	Autoencoders can reduce the dimensionality of data, making it more manageable and easier to process.	Autoencoders may struggle with very complex data, as they might not capture intricate patterns effectively.	[69,118,92-94]
RNN	An RNN, is a particular sort of neural network for coping with time series data or sequences.	RNNs are specifically designed to process sequences of data, making them ideal for tasks like time series analysis, natural language processing (NLP), and speech recognition	RNNs are prone to vanishing and exploding gradient issues during training, which can hinder learning and make it difficult to capture long-term dependencies	[42,88,41,60, 89]
RvNN	A particular kind of neural network called a recursive neural network (RvNN) is made to deal with hierarchical data structures like parse trees in natural language processing.	RvNNs are perfect for jobs like syntactic parsing and sentiment analysis because of their exceptional efficacy in processing hierarchical data structures.	Training RvNNs can be complex and computationally intensive, especially for large datasets or very deep hierarchical structures	[90,91]
CNN	The CNN recognizes larger chunks of the input with each layer. As the data moves through the layers, it begins to identify greater parts of the object, with the earlier levels concentrating on simpler traits.	CNNs are highly efficient at processing images due to their ability to apply convolutional filters that capture spatial hierarchies in the data	Training CNNs can be computationally intensive, requiring significant processing power and memory, especially for large models and datasets	[8,84,85,87,42, 84,85]
GRU	Like LSTM, GRU is a gating mechanism used in recurrent neural networks; however, unlike LSTM, it does not include an output gate. It has been shown that GRUs perform better on certain smaller and less common datasets.	GRUs have a simpler structure compared to LSTMs, with fewer gates (reset and update gates).	While GRUs handle long-term dependencies better than traditional RNNs, they may not capture them as effectively as LSTMs in some complex tasks.	[119–122,111, 89]
GAN	GAN is used for unsupervised learning. GANs consist of two neural networks: a generator and a discriminator. The generator creates artificial data that resembles actual data, while the discriminator distinguishes between real and fake data.	GANs are capable of generating highly realistic data, such as images, videos, and audio, which can be nearly indistinguishable from real data.	GANs can be difficult to train due to the delicate balance required between the generator and discriminator.	[95,96,115,123]
Transformer	Transformers, particularly in the context of neural networks, have revolutionized various fields, especially NLP.	Transformers can process entire sequences of data simultaneously, unlike RNNs which process data sequentially.	Transformers demand substantial computational resources, which include high-performance GPUs and significant memory capacity, particularly when it comes to training large models.	[97,124,98, 99–101,102, 125,65, 126–128,32]
Boltzmann Machines	They are a type of stochastic network that can learn complex distributions over their inputs.	Boltzmann Machines can learn from unlabeled data, making them useful for unsupervised learning tasks such as feature extraction and dimensionality reduction.	Training Boltzmann Machines is computationally intensive and slow, especially for networks with many layers or units.	[58,106,85,59, 110,63,95]
Ensemble	This involves fusing the benefits of the individual DL models to enhance the general performance of the final model.	Ensemble methods often achieve higher accuracy than individual models.	Ensemble models are more complex to build and interpret compared to single models.	[113,9,33,129, 113,116,3,115, 117]

model incorporating attention mechanisms. The attention method enhances the overall effectiveness of FND by helping to better capture the context. In 2023, Zhang [91] introduced a novel model called depth-breadth tree-structured recursive neural networks. The global propagation information may be obtained by using the model propagation tree to extract the depth propagation features of the multi-branch and the breadth features at the same depth. The authors use an attention method during the extraction of depth propagation characteristics to modify the significance of the hints, enabling the model to concentrate on the crucial clues that facilitate the identification of fake news. One-dimensional convolutional network is used to extract breadth propagation features, first obtaining fine-grained relationships among word levels within the comments and then, in a subsequent step, obtaining semantic interaction between the comments, to supplement the depth propagation features. Results from Twitter dataset showed that the suggested technique is more accurate in detecting early false news than the current unidirectional fake news detection network in the propagation tree.

4.4. Auto encoder (AE)

A few emphases of the autoencoder model have been grown as of late to classify multimodal news things as genuine or deceitful. A text and visual extractor are a part of both the encoder and the decoder. To detect fake news, for example, [69] introduced a start-to-finish network called multi-modal variational autoencoder (MVAE), which consolidates a twofold classifier with a bimodal variational autoencoder. Considering this, a multimodal portrayal procedure (MVAE-FakeNews) was introduced by [92] to distinguish sham news in a solitary class mark. The recommended technique joins two modalities, message embeddings and subject data that are believed to be promising for misleading news identification to show a Multimodal Variational Autoencoder another portrayal. The creators utilized three datasets in their examination, considering the Portuguese and English dialects. Reference [92] proposed a model for fake news detection that encapsulates texts using a single class label, which learns a novel representation by integrating various promising modalities for news data, including text embeddings, topics, and linguistic information. A multimodal-based variational

autoencoder model for false news detection was described by [93]. To improve discriminatory power, the suggested model mixes the data from the text and picture modalities. The intended model is known as a multi-model variational autoencoder since it uses both text and picture data from false news to extract textual and visual news features and process both features concurrently into a variational autoencoder. To be more precise, BERT and VGG19 embeddings are created for the text and picture modalities, respectively. Furthermore, a unique method for detecting false news is investigated by [94] using a Gated Variational AutoEncoder (GatedVAE). This technique tackles the problem of noise in multimodal data, which may cause misclassifications in current algorithms. An example of this kind of noise is irrelevant information seen in news captions. The GatedVAE incorporates a gating mechanism to suppress noisy inputs, dynamically controlling the passage of information and enhancing the model's ability to learn effective multimodal representations.

4.5. Generative adversarial network (GAN)

The unsupervised DL method class known as GAN is composed of a discriminator that aims to categorize the sample as real or fabricated and a generator that learns to create new example data with the same features as training data. Every epoch, the generator is modified to produce more credible examples more efficiently, while the discriminator is updated to improve its ability to categorize the samples. It is used commonly for altering photos and making DeepFakes. Using adversarial networks and similarity inference, [95] provide a novel method of combating false news on social media platforms to identify deepfake. For textual feature extraction, the proposed model combines the Bidirectional Encoder Representations from Transformers and CNN models; for visual feature extraction, the VGG-19 model is used. By using similarity learning and reasoning, this model creates representations of similarities between textual and visual characteristics, which improves detection accuracy. To further increase the robustness of the model, adversarial networks are used to investigate the connection between false news and events. In addition, [96] introduced a framework for identifying multi-modal deceptive contents. It enhances the accurate localization of manipulated regions within both images and textual annotations using a GAN-based approach. The model utilizes a pretraining approach that combines visual and language data to enhance the detection of deepfakes. This pretraining helps the model understand the intricate relationships between images and their corresponding textual descriptions. The VLP-GF model demonstrates superior performance in identifying manipulated content across various datasets. It effectively localizes manipulated regions within images and textual annotations, making it a robust tool for MFND.

4.6. Transformer

Transformer-based network has been around for some time and have demonstrated to be very successful for an assortment of regular language handling applications. It oversees long-range conditions and endeavors to tackle succession to grouping difficulties. It doesn't utilize arrangement-adjusted RNNs or CNNs; all things considered; it utilizes self-regard to produce portrayals of its contributions as well as its outcomes. A portion of the habitually used transformer-based network that has been sent really in misleading news discovery are BERT, RoBERT, and ALBERT [97–104]. For instance, [97] presented an educational program based on multi-modular veiled transformer organization (CMMTN) intended to demonstrate both between and intra-methodology cooperations of multi-modular data while sifting through insignificant settings during the cycles to identify positive unlabeled multi-modular fakenews. In particular, the model uses BERT and ResNet to create further developed portrayals of text and pictures, separately. To handle the multimodal false news detection, [66] suggested a multi-modal fusion and transformer-based model. In particular, the model uses several transformers to extract text and visual information, then attention processes combine those features. Additionally, it makes use of the classifier's ability to distinguish between text and picture features to enhance the model's performance. The efficacy of the suggested model is shown by the experimental findings on publicly available datasets. Reference [98] presented a multi-modal transformer that utilizes two levels of visual features for FND. Initially, the method consistently represents news text and images as sequences that can be processed by a transformer. To enhance the utilization of news images, the approach incorporates two types of visual features: global features and entity-level features. A picture inscription-based methodology was given by [99] to build the model's ability to remove semantic data from photos. Officially, the methodology consolidates realities about picture depictions into the text, spanning the significance hole between language and visuals. Reference [100] proposed a likeness and transformer-based acknowledgment framework named misleading News Revealer (MNR), which joins texts and photographs of information to distinguish false news.

4.7. Boltzmann machine

This is a sort of neural network with recurrent neurons in which the nodes have biases and make binary decisions. More complicated systems, like a Deep Boltzmann Machine (DBM), may be created by stacking several Boltzmann Machines (BM). Compared to BM, these networks include more hidden layers and random connections between their nodes. Determining the true relationship between these two sets of variables is the training challenge. A BM uses a Markov random field for layer-wise pre-training on the big unlabeled dataset before giving older layers feedback. While the concept of the RBM is similar to that of the Encoder-Decoder, it employs stochastic units with a specific distribution rather than a deterministic one. When it comes to feature selection and extraction, RBM is crucial for dimensionality reduction, classification, regression, and many other processes. In addition, [105] offers a multimodal DL model for the FND that is based on the Deep Boltzmann Machine.

4.8. Attention mechanism

By concentrating on significant input components, attention mechanisms improve the computational efficiency and prediction accuracy of DL models [58,106–109]. A multi-modal approach for identifying false news was presented in [58]. It utilizes suitable multi-modal feature fusion to optimize the relationship between text and picture data, resulting in an efficient multimodal shared representation. Experimentally, they exhibit how integrating text and image might improve performances. The proposed approach utilizes Multimodal Factorized Bilinear Pooling (MFB) for visual element extraction. To assess the model's presentation, tests were led on two freely accessible datasets, Weibo and Twitter. The outcomes demonstrate that this technique altogether outflanks existing high-level models. Furthermore, [106] used Repetitive Brain Organizations with a consideration component (att-RNN) to investigate and coordinate multimodal data. The consideration instrument improves the model's capacity to distinguish counterfeit news by permitting it to zero in on the most relevant parts of the information. Reference [59] developed a unique Multimodal Co-Consideration Organizations (MCAN) to all the more likely incorporate etymological and visual data for counterfeit news identification, provoked by the manner in which individuals consume news blending picture and text. Extensive analyses on two genuine world datasets uncover that MCAN outflanks best in class draws near and has the ability to learn between conditions across multimodal highlights. Multimodal coordinating mindful co-consideration networks with shared information refining were proposed by [110] to improve the identification of sham news. To improve multimodal fusion, an image-text matching-aware method was specifically created to learn the alignment of the

picture and texts. One may acquire the image-text matching representation by using a pre-trained vision-language model.

4.9. Gated recurrent units (GRU)

The GRU model attempts to overcome the vanishing gradient problem of a regular recurrent neural network [111,89,112]. GRU may also be seen as a variant on the LSTM since both are built similarly and, in certain situations, give equally outstanding results. To do this, [111] presented a technique that combines textual and visual data to identify fake news. First, the model employs a two-branch learning strategy to extract more useful characteristics from the modality's hidden layer data. Subsequently, they suggested an attention technique to capture multimodal internal linkages to detect fake news, as well as a multimodal bilinear pooling mechanism to better combine textual and visual characteristics. The experimental results demonstrated that the proposed method outperformed the current state-of-the-art method on publicly accessible Weibo and Twitter datasets. In addition, [89] present an innovative approach to tackling fake news by leveraging both intra-modality and inter-modality features. The model extracts and aggregates features within each modality such as text and image to reduce noise and redundancy. This step ensures that the most relevant features are highlighted. The model was tested on Weibo and Twitter datasets, demonstrating superior performance compared to other advanced methods.

4.10. Ensemble

Ensemble method combines multiple models to improve overall performances. By leveraging the strengths of different algorithms, they can achieve higher accuracy and robustness compared to individual models [113,114,82]. Reference [115] have presented a solo technique given Autoencoder and GAN for multimodal fake news identification, driven by the viability of troupe draws near. The creators made the high layered include a vector of information phrases with the utilization of an autoencoder. Generators in GAN then, at that point, utilize this vector to deliver machine-created misleading news. A discriminator is then used to recognize the fake news that was made and the genuine news. Around 30,000 news titles are remembered for the news dataset, which is utilized to test the recommended model's presentation. The exploratory outcomes recommend that the proposed strategy is much more solid and can be exceptionally useful in computerizing counterfeit news recognition. A multi-modular misleading news discovery strategy given outfit learning was proposed by [113]. It first takes utilization of an openly available dataset called Fakeddit, which has north of 1000,000 instances of misleading news. Then, it preprocesses news text-based content utilizing Regular Language Handling. Then, it utilizes Visual Bidirectional Encoder Portrayals from Transformers (V-BERT) to make embeddings for the text and image of the connected news. An ensemble of deep learners based on attention processes was used by [33] to improve the efficacy of the FND technique. An ensemble model's ability to succeed is largely dependent on how diverse its students are. To do this, the authors developed a unique loss function that, on the one hand, compels learners to focus on various sections of news material and, on the other, produces high classification accuracy. Similarly, [116] introduced a multimodal system for identifying misleading news that utilizes visual picture credits through scientific investigation and picture inscribing, as well as concealed design extraction capacities from text utilizing various leveled Consideration Organization (HAN). The obtained findings further demonstrate that the combined model outperforms individual techniques in properly identifying false information, this is also in line with [117] study, where Ensemble graph neural networks was adopted to detect fake news using user engagement and text features.. Additionally, a quantum multimodal fusion-based model for false news detection (QMFND) was presented by [3]. To achieve discriminative findings, QMFND combines the retrieved pictures and textual data and then feeds them through a proposed quantum convolutional neural network.

Table 4 presents a detailed comparison of different deep learning-based approaches used for multimodal false news detection (MFND). It provides various models, describing their fundamental functionalities, strengths, and weaknesses, along with corresponding references. The table covers models including Attention Mechanisms, Autoencoders (AE), Recurrent Neural Networks (RNN), Recursive Neural Networks (RvNN), Convolutional Neural Networks (CNN), Gated Recurrent Units (GRU), Generative Adversarial Networks (GAN), Transformers, Boltzmann Machines, and Ensemble methods. Every model has various applications: Attention Mechanisms, for instance, reinforce attention to required data elements, while RNNs and GRUs are utilized for sequential data processing. CNNs are suitable for image-based processing, and Transformers revolutionized NLP by processing entire sequences parallelly. The table highlights both the strengths of these models—e.g., GANs' ability to create realistic synthetic data and ensemble methods' improved accuracy—and their weaknesses, including high computational requirements and training challenges.

Closer inspection reveals model trade-offs. Some models, like Attention Mechanisms and Transformers, enhance contextual understanding at the expense of high computational power, thus making them less suitable for low-power consumption. Similarly, CNNs extract spatial hierarchies well but require high processing capacity. Autoencoders and Boltzmann Machines work well for dimensionality reduction but poorly with complex patterns. RvNNs are particularly suited for hierarchical data structures but are computationally intensive to train. Ensemble methods, while they have an accuracy edge, add another layer of complexity, making interpretation and implementation more challenging. Overall, the table presents a balanced trade-off of the strengths and weaknesses of deep learning models in MFND to guide researchers in selecting appropriate models for the specific requirements of their applications.

5. Standard datasets used to evaluate the performance of DL models for the FND

To answer this question, we present an overview of the evaluation data sets utilized for multimodal fake news detection, categorizing them based on the modalities included.

5.1. Text and image dataset

The following data sets for fake news detection comprise of image and text modalities:

Twitter dataset: The evaluation datasets for the Multimedia Use task in the Media-Eval 2015 and 2016 workshops were the image-verification corpus datasets for Media-Eval-2015 and Media-Eval2016 [58,41], commonly referred to as Twitter datasets. The training set for the Media-Eval-2015 dataset consists of 11 events, totaling 5008 actual and 6840 fraudulent tweets. There are 2564 fake tweets and 1217 actual ones in the test batch. There are 192 more fake tweets in MediaEval2015 than there were in previous years because certain rumor tweets were included but later removed from the final dataset. Regarding MediaEval2016, a fresh collection of 1107 (actual) and 1121 (false) posts was used for testing, and the training and test sets from Media-Eval-2015 were blended into one set.

Weibo dataset: The Weibo dataset for fake news detection [41,55] was collected from various sources and verified by Weibo's rumor-debunking platform. The real news in this dataset was compiled from trustworthy Chinese media outlets, such as Xinhua News Agency, between May 2012 and January 2016. The original dataset contains 9528 news items, including 4749 false news and 4779 true news, all of which include images. In the authors' experiment, the training set involves 3749 fake news and 3783 true news,

while the test set includes 1000 fake and 996 true news items. There are multiple versions of the Weibo dataset based on the time frame of data gathering; for example, [41] utilized data from May 2012 to November 2018.

PolitiFact: The PolitiFact dataset is a component of the news data repository FakeNewsNet [65], which verifies the accuracy of political reporting and concerns. It was gathered from the organization's fifth website. Three contexts make up this information: the news content, the social context, and the spatiotemporal information. As part of the dataset generation process, human annotators have labeled categories. The news headline and body make up the majority of the news content component. The news stories on PolitiFact were released between May 2002 and July 2018. There are 624 true news stories and 432 false news pieces out of 1056 total. Throughout the whole story, there are 948 occurrences with textual information; 528 of them are genuine and 420 are false. 783 news stories have visual (picture) content; 336 of them are false, and 447 are true.

GossipCop (GCop): FakeNewsNet also includes the GossipCop dataset [42]. Gossipcop is a website that concentrates on entertainment and celebrity news, but it also fact-checks news items. This dataset contains news stories that were released between July 2000 and December 2018. In regards to annotation, contexts, and contents, it is similar to PolitiFact. 22,140 news items total on Gossipcop, including 5323 fraudulent news and 16,817 true news. There are 21, 641 news items containing textual content; 4947 of them are fraudulent, while 16,694 are authentic. There are 18,417 occurrences with visual contents, 16,767 of which are authentic and 1650 of which are fraudulent. According to these statistics, GossipCop is an unbalanced dataset by class.

Fakeddit: 1063,106 news occurrences across many categories may be found on Fakeddit [65,70]. It's among the biggest multimedia compilations of false information. The contents also include meta-data and comments in addition to the text and picture modes. A comprehensive annotation process has been used to arrange Fakeddit into precise binary, 3-way, and 6-way categories.

5.2. Text and video dataset

The following experimental datasets for the identification of false news mostly consist of text and video modalities:

VAVD, or the Volunteer Annotated Video Dataset: As the name implies, 20 volunteers' efforts resulted in the creation of VAVD [46]. Between September 2013 and October 2016, over 100,000 videos and comments were posted to YouTube. These were gathered and categorized. Annotations were divided into three categories, Legitimate Spam, Not Sure, and Others, using a two-round annotation procedure.

Fake Video Corpus (FVC): As part of the InVID project, FVC [46] was created. It is made up of videos together with associated information. Due to its ongoing growth, it has many versions; the most recent version has 2458 actual and 3957 fraudulent videos.

The Misleading Youtube Video Corpus (MYVC) was created by gathering both authentic and fraudulent news for well-known fact-checking websites and YouTube [43]. There are 902 fake and 903 actual news items in it. The authors, [43] used this dataset with FVC and VAVD in their experiment.

5.3. Audio and video dataset

The following datasets, which are based on audio and video modalities, are briefly described:

Deepfake Detection Challenge (DFDC): The Deepfake Detection Challenge entries were assessed using the DFDC dataset [48]. Direct video recording of participants was used to get the raw data. Training, validation, and test sets were created from the final dataset. The training set is made up of 119,154 10-second video clips with 486 distinct individuals, 100,000 of which are deepfakes. The validation set consists of four thousand ten-second video clips, 2000 of which are deepfakes spanning 214 distinct topics. Five thousand of the ten-second video clips in the test set are deepfakes.

DeepFake-TIMIT, the Audio-Video TIMIT dataset: 43 people's short-sentence videos and matching audio recordings are included in the TIMIT collection [48]. The data was gathered throughout three sessions, with an average gap of seven days between the first and second sessions and six days between the second and third. The sentences were selected from the TIMIT corpus test section. Each individual has 10 sentences. Session 1 covers the first six sentences. Session 2 will cover the following two sentences, and Session 3 will cover the last two. The first two phrases are the same for every person, but the next eight are often unique to each individual.

FakeAVCeleb (FkAVCD): The 20,000 movies in the FakeAVCeleb dataset [49] are made up of 500 actual and 19,500 fraudulent films. There's audio to go along with every video. One notable feature of this dataset is its mixing of actual and false modalities, which makes it suitable for a wide range of classification problems.

World Leaders Deepfake Dataset (WLDD): Focuses on world leaders who are each referred to as persons of interest (POI) [78]. We obtained the uncut videos straight from YouTube. The whole dataset is divided into sections, with the actual portion consisting of 30,683 ten-second films featuring 1004 distinct individuals. The fake section includes lip synchronization deep fakes, face-swap deep fakes, puppet master deep fakes, and humorous impersonators for every point of interest.

Presidential Deepfakes Dataset (PDD): This dataset [78] includes 32 films featuring the two most recent US presidents, Donald Trump and Joe Biden. In every video in the dataset, one of the two presidents is shown in an official setting discussing politics. To produce the deepfakes, 16 out of the 32 had their contents altered. The audio and visual (video) modalities are among the changed characteristics.

5.4. Text, audio and video datasets

FakeSV Dataset: The 1827 false and 1827 actual news examples in the FakeSV Dataset [44] were gathered from Kuaishou and Douyin. Chinese applications Douyin and Kuaishou allow users to share their short films. Every news instance includes a user, a title, meta-data, and a video with audio.

Table 5 shows the summary of all the datasets used by the selected articles for multimodal fake news detection.

Social media datasets like Twitter and Weibo are mainly used for sentiment analysis, trend analysis, and disinformation studies. They suffer from some severe limitations, particularly cultural and linguistic biases. Twitter has a largely English-speaking client base, while Weibo has a Chinese-speaking client base, making comparisons challenging. Linguistic idiosyncrasies, idiomatic usage, and cultural context influence user expression, thereby impacting sentiment analysis and text interpretation. Furthermore, Weibo is subject to stringent governmental censorship, prompting self-censorship and content moderation practices that are quite disparate from the relatively freer discourse climate characteristic of Twitter. All of these conditions result in implicit biases in data representation and limit the generalizability of conclusions across platforms.

Apart from linguistic and cultural dissimilarities, other concerns include demographic disparities, content formats specific to each platform, and usage patterns. Twitter users are scattered globally, while Weibo users are largely concentrated in China, making it difficult to apply findings from one platform to the other. Differences in post length, engagement measures, and algorithmic visibility of content further complicate cross-platform analysis. Furthermore, data gathering is often restricted by the limitations of APIs, sampling biases, and the prevalence

Table 5
Description of the datasets used for MFD models.

Datasets	Year of release	Statistics	Constituent	Labels	source	Used in
Twitter dataset [41]	2015	7032 (F) 361 (I) 5008 (R)	Text, visual	2	Twitter	[41,58,59,38,64]
Weibo dataset [41]	2016	4749 (F) 9528 (I) 4779 (R)	Text, visual	2	Twitter	[41,55,58,59,62,38,63,64]
PolitiFact dataset [65]	2018	4749 (F) 9528 (I) 4779 (R)	Text, visual	2	PolitiFact.com	[65,42,65,53,81]
Gossipcop dataset [65]	2018	9528 (I) 4749 (F) 4779 (R)	Text, visual	2	GossipCop.com	[42,57,65]
Volunteer Annotated Video Dataset (VAVD) [46]	2016	100,000 (V)	Text, video	3	YouTube	[46]
Fake Video Corpus (FVC) [46]		6515 (V) 2458 (R) 3957 (F)	Text, video	2	Multimedia Knowledge and Social Media Analytics Lab (MKLab) at the Information Technologies Institute (ITI).	[46]
Misleading Youtube Video Corpus (MYVC) [71]		1805 (V) 902 (F) 903 (R)	Text, video	2	YouTube	[43,71,47]
Deepfake Detection Challenge (DFDC) [48]	2015	119,154 (V) 100,000 (F) 486 (R)	Audio, Video	2	YouTube videos	[48,75,77]
FakeAVCeleb [49]		500 (R) 19,500 (F) 20,000 (V)	Audio, Video	4	YouTube videos	[49,72,73,76–80]
FakeSV [44]		1827 (F) 1827 (R)	Text, audio and video	2	YouTube videos	[44]
FakeNewsNet [66]	2019	5367(F) 19,200 (I) 17,222 (R)	Text, visual	2	PolitiFact, GossipCop,	[130]
PHEME [63]	2016	1972 (F) 2672 (I) 3830 (R)	Tweet, conversational threads	3	Twitter	[63]

Note: I—Total Number of Images, V—Total Number of Videos, R—Number of Real news, F—Number of Fake news.

of automated bots, all of which can distort research findings. Ethical challenges, such as user privacy, content moderation, and censorship-induced data gaps, also influence the validity of research drawn from these data. Keeping these constraints in consideration, researchers will have to take great care to account for these biases when they make sense of findings derived from Twitter and Weibo data

6. Evaluation measures for fake news detection models using deep learning

Evaluation metrics must be carefully considered when assessing a model's efficacy. While a model may attain more classification accuracy while being built, it is crucial to ascertain whether it can successfully solve a particular issue in various settings. Making this determination merely based on categorization accuracy is frequently insufficient. For a thorough examination, more performance metrics are required. It is more difficult to develop a promising strategy than it is to create a model since a prospective technique must pass the review of performance measures. Several assessment indicators are used to assess a model's effectiveness. A vital tool that helps arrange and structure an assessment is the confusion matrix. An effective method for evaluating a model's performance on a test dataset is a confusion matrix. Along with a summary of the performance of the model, it provides useful discoveries such as true positives, true negatives, false positives, and false negatives. Traditional measures of model performance that have been examined in some research include accuracy, precision, recall, F-score measure, area under the curve (AUC), false-negative rate (FNR), and Cross Entropy Loss. Table 6 shows some of the evaluation metrics used based on the publications selected for this study.

7. Limitations of the study

This SLR identifies numerous deep learning-based models for fake news detection. By creating research protocols, we hope to maximize both internal and external validity while responding to the RQs. The following list of current problems and restrictions that this argument must overcome is provided in this section:

- This SLR only includes publications from journals and conferences that explain deep learning-based false news detection methods. We discovered some irrelevant research publications early on in the study as a result of our search strategy, which we later removed from our review. The chosen research articles are guaranteed to satisfy the study's needs in this way. It is thought that including more sources like additional sourcebooks, for example, would have enhanced the review.
- We restricted our search to works that were solely written in English. Since related articles on this topic can be available in other languages, this could lead to linguistic bias. Thankfully, all of the sources we used for this research were written in English.
- Even though the principal databases were taken into consideration when analyzing the study articles, it is possible that other digital libraries having pertinent studies were overlooked. We evaluated the keywords and searched queries to a renowned collection of research works to overcome this restriction. However, certain synonyms may be overlooked when searching for the keywords. To overcome this problem and ensure that no crucial words are missed, the SLR protocol has been revised.

8. Emerging trends and future directions

This section covers several challenging fake news detection issues.

Table 6
List of some common metrics used for evaluation.

Metric used	Formula	Description	References
Accuracy	$Accuracy = \frac{TruePositive + TrueNegative}{TotalNumberofPrediction}$	The accuracy score, often referred to as the classification accuracy rating, is calculated as the ratio of correct predictions to all predictions made by the model.	[38,41,42, 55,40, 56–60,65, 67,47,48, 75,78,79, 131]
Precision	$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$	Precision is determined as the proportion of true positives discovered to all positive results, including those that were mistakenly identified.	[79,59,62, 63,49,74, 53]
Recall	$Recall = \frac{TruePositive}{TruePositive + FalseNegative}$	How successfully a classification model can locate all the pertinent occurrences is gauged by recall.	[132,59, 62,63,70, 49,74,76, 79,53]
F1-Score	$F1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$	For each class, the model's accuracy is determined by the F1 score. The F1-score measurement is frequently applied in cases where the dataset is imbalanced	[79,81,64, 66,45,71, 77]
Cross Entropy Loss	$Cross\ Entropy\ Loss = -(y\log(p) + (1-y)\log(1-p))$	Cross Entropy Loss is a measure of the difference between two probability distributions for a given set of events or observations.	[35]
AUC- ROC Curve	$FPR = \frac{FalsePositive}{FalsePositive + TrueNegative}$	This curve uses a True Positive Rate (Recall) and False Positive Rate (FPR). The AUC metric evaluates the full two-dimensional field within the context of the entire ROC curve	[43]
Computational Cost	$Computational\ Cost = FLOPs \times Time \times Energy$	The computational cost can be quantitatively measured using metrics such as Floating Point Operations Per Second (FLOPS), which indicate the number of operations a model performs.	[133]

Early detection of fake news is especially crucial for preventing its continued propagation. It becomes crucial to investigate the issue in a partially supervised situation, that is, with few or no labeling for training, because obtaining the ground truth and labeling the false news dataset are labor- and time-intensive tasks. Furthermore, the multi-modality of fake news on social media has received comparatively little research. Recognizing the rationale behind a news article's categorization as bogus by machine learning models is also crucial since it might yield additional information and insights that are obscured by content-based algorithms.

The detection of fake news has multiple drawbacks. The following research gaps have been discovered following a thorough analysis and assessment of the selected literature that has been reviewed in Section 2.

i. Multimodality: Prior to the rise of social media, news stories were only text-based; today, they are multimodal. Videos and pictures are now incorporated in addition to the texts in news contents. As such, creating a method that can function on a multimodal dataset is imperative. Though numerous methods have been created to address this issue, the multimodal component of false news on has not received much attention. This is because handing multimodal setting can be challenging on its own. While some of the proposed methods have proven useful in detecting misinformation, they are still unable to determine the relationships between different modalities.
ii. Multilinguality: Developing models for multilingual and cross-lingual fake news detection has received comparatively less attention than the majority of research conducted thus far in this topic, which focuses on the identification of fake news in the English language. Language usage is unrestricted since social media platforms are used by a significant section of the world's population. Additionally, gathering fake articles and its annotation in other languages is very time-consuming and labour-intensive.
iii. Varying degrees of falseness: most of methods now in use for identifying false news take a binary approach to the problem. A news item may, in fact, include both accurate and false information. It's crucial to classify fake news based on how misleading it is as a result. However, since the boundaries between classes are more complex, the classifier must provide better discriminative capability and more robustness as the number of categories increases for multiclass fake news detection.
iv. Early detection: Fake news have seriously harm people and society at large. The potential for false news to harm someone's or an organization's reputation makes it imperative to recognize it as soon as possible. Fake news early detection aims to halt the spread of false information by offering notifications in advance.
v. Insufficient datasets: An additional problem is the data source: there aren't many databases on false news, and the public ones are rather small. There is a lack of multi-modal datasets in this issue space. Table 5 in the section above compares the primary benchmark multimodal fake news datasets. The algorithm's performance could be negatively impacted by the abundance of garbage values and irrelevant information found in unstructured data.

9. Conclusion

With an emphasis on multimodal context, we have provided description of current existing methods and strategies for addressing the problem of identifying false news on social networks in this study. We have focused our study mainly on five important areas. Initial, an exact meaning of fakenews is given in the review, alongside a characterization of related phrasing and a very much determined scientific classification of misleading news recognition strategies. Through the review, it was found that the multimodal part of information content has been generally neglected. Moreover, a few generally utilized profound learning models, structures, libraries, and move learning methods generally remarkably TensorFlow, have been underscored. Besides, taking into account the utilization of multimodal information, we have given an outline of cutting edge approaches for distinguishing counterfeit news via virtual entertainment stages utilizing profound learning methods. The study demonstrates that RNN-based methods are utilized to preserve the sequential data included in the textual information, whereas CNN-based models are often employed to handle picture data. Furthermore, several attention network adjustments are used in order to maintain the association between the picture and text. These models lack the ability to analyze multilingual data, which is common when using social media, and instead utilize English as their main language for detection. Fourth, the study clarifies different methods for extracting data and sources from it. For this specific endeavor, only a small number of multimodal datasets are accessible since this area of study is still relatively new. In conclusion, future directions of research in the area of fake news detection on social networks should focus on developing multimodal approaches that leverage different data modalities, such as text, images, and videos. The design of multilingual detection systems is necessary, considering the current approaches mostly rely on the English language. There is also a need for greater availability of datasets for allowing better model training and testing. The exploration of new deep learning methods, such as hybrid models and transformer models, is needed to enhance detection accuracy. Furthermore, maintaining the relationships among various modalities, facilitating data extraction techniques, and solving the problems of multimodal data management are also crucial. Ultimately, the facilitation of real-time fake detection models and integration of explainable AI (XAI) to multiple social media platforms will be necessary for effective practical implementation.

CRediT authorship contribution statement

Maged Nasser: Writing – original draft, Visualization, Validation, Project administration, Methodology, Data curation, Conceptualization. **Noreen Izza Arshad:** Writing – original draft, Validation. **Abdulalem Ali:** Resources, Methodology, Formal analysis, Data curation. **Hitham Alhussian:** Writing – review & editing, Validation, Methodology. **Faisal Saeed:** Writing – review & editing, Validation, Formal analysis. **Aminu Da'u:** Writing – review & editing, Visualization, Resources. **Ibtehal Nafea:** Formal analysis, Writing – review & editing.

Declaration of competing interest

There is no conflicts of interest

Acknowledgments

The authors extend their appreciation to the Universiti Teknologi PETRONAS for funding this research through a Short-Term Internal Research Funding (STIRF) Grant (Grant Number: 015LA0-057) and Yayasan Universiti Teknologi Petronas (YUTP) 015LC0-483

Data availability

No data was used for the research described in the article.

References

[1] R.K. Kaliyar, A. Goswami, P. Narang, A hybrid model for effective fake news detection with a novel COVID-19 dataset, ICAART (2) (2021) 1066–1072.
[2] A.B. Athira, S.D.M. Kumar, A.M. Chacko, A systematic survey on explainable AI applied to fake news detection, Eng. Appl. Artif. Intell. (2023).
[3] Z. Qu, Y. Meng, G. Muhammad, P. Tiwari, QMFND: a quantum multimodal fusion-based fake news detection model for social media, Inf. Fusion (2024), https://doi.org/10.1016/j.inffus.2023.102172.
[4] M.F. Mridha, A.J. Keya, M.A. Hamid, et al., A comprehensive review on fake news detection with deep learning, IEEe Access. (2021), https://doi.org/10.1109/ACCESS.2021.3129329.
[5] Q. Chang, X. Li, Z. Duan, Graph global attention network with memory: a deep learning approach for fake news detection, Neural Netw. 172 (2024) 106115.
[6] Y.M. Rocha, G.A. de Moura, G.A. Desidério, et al., The impact of fake news on social media and its influence on health during the COVID-19 pandemic: a systematic review, J. Public Heal. (2023).
[7] M.T. Zamir, F. Ullah, R. Tariq, et al., Machine and deep learning algorithms for sentiment analysis during COVID-19: a vision to create fake news resistant society, PLoS. One 19 (2024) e0315407.
[8] I. Segura-Bedmar, S. Alonso-Bartolome, Multimodal fake news detection, Information (2022), https://doi.org/10.3390/info13060284.
[9] P.K. Verma, P. Agrawal, V. Madaan, R. Prodan, MCred: multi-modal message credibility for fake news detection using BERT and CNN, J. Ambient Intell. Humaniz. Comput. (2023), https://doi.org/10.1007/s12652-022-04338-2.
[10] A. Khraisat, C.L. Manisha, J. Abawajy, Survey on deep learning for misinformation detection: adapting to recent events, multilingual challenges, and future visions, Soc. Sci. Comput. Rev. (2025) 08944393251315910.
[11] K. Shu, H. Liu, Detecting fake news on social Media. Synthesis Lectures on Data Mining and Knowledge Discovery, 2019, https://doi.org/10.2200/s00926ed1v01y201906dmk018.
[12] M.J. Page, J.E. McKenzie, P.M. Bossuyt, et al., The PRISMA 2020 statement: an updated guideline for reporting systematic reviews, BMJ 372 (2021), https://doi.org/10.1136/bmj.n71.
[13] B. Kitchenham, S. Charters, Guidelines for performing systematic literature, Rev. Softw. Eng. Eng. 65 (2007), https://doi.org/10.1145/1134285.1134500.
[14] N. Genc-Nayebi, A. Abran, A systematic literature review: opinion mining studies from mobile app store user reviews, J. Syst. Softw. 125 (2017) 207–219.
[15] B. Kitchenham, O. Pearl Brereton, D. Budgen, et al., Systematic literature reviews in software engineering - A systematic literature review, Inf. Softw. Technol. 51 (2009) 7–15, https://doi.org/10.1016/j.infsof.2008.09.009.
[16] C. Comito, L. Caroprese, E. Zumpano, Multimodal fake news detection on social media: a survey of deep learning techniques, Soc. Netw. Anal. Min. (2023), https://doi.org/10.1007/s13278-023-01104-w.
[17] E. Hashmi, S.Y. Yayilgan, M.M. Yamin, et al., Advancing fake news detection: hybrid deep learning with fasttext and explainable ai, IEEE Access (2024).
[18] H.T. Phan, N.T. Nguyen, A dual LSTM-based multimodal method for fake news detection, in: European Conference on Artificial Intelligence, Springer, 2024, pp. 3–14.
[19] D.T.T. Thuy, L.T.M. Thuy, N.C. Bach, et al., Designing a deep learning-based application for detecting fake online reviews, Eng. Appl. Artif. Intell. 134 (2024) 108708.
[20] K.M. Karaoğlan, Novel approaches for fake news detection based on attention-based deep multiple-instance learning using contextualized neural language models, Neurocomputing 602 (2024) 128263.
[21] J. Jouhar, A. Pratap, N. Tijo, M. Mony, Fake news detection using python and machine learning, Procedia Comput. Sci. 233 (2024) 763–771.
[22] C. Liu, X. Wu, M. Yu, et al., A two-stage model based on BERT for short fake news detection, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2019.
[23] H. Jwa, D. Oh, K. Park, et al., exBAKE: automatic fake news detection model based on bidirectional encoder representations from transformers (BERT), Appl. Sci. (2019), https://doi.org/10.3390/app9194062.
[24] A.S. Arun, V.P.G. Subhash, S. Shridevi, Fake news detection in mainstream media using BERT. Lecture Notes on Data Engineering and Communications Technologies, 2023.
[25] S.K. Uppada, B.S. Ashwin, B. Sivaselvan, A novel evolutionary approach-based multimodal model to detect fake news in OSNs using text and metadata, J. Supercomput. (2024), https://doi.org/10.1007/s11227-023-05531-6.
[26] K. Shu, L. Cui, S. Wang, et al., Defend: explainable fake news detection, in: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2019.
[27] R. Tan, B.A. Plummer, K. Saenko, Detecting cross-modal inconsistency to defend against neural fake news, in: EMNLP 2020 - 2020 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference, 2020.
[28] H. Chen, H. Wang, Z. Liu, et al., Multi-modal robustness fake news detection with cross-modal and propagation network contrastive learning, Knowledge-Based Syst. 309 (2025) 112800.
[29] A. Wani, I. Joshi, S. Khandve, et al., Evaluating deep learning approaches for Covid19 fake news detection, in: Communications in Computer and Information Science, 2021.
[30] Z. Pehlivan, On the pursuit of fake news: graph Neural Network meets NLP, in: CEUR Workshop Proceedings, 2021.
[31] A. Mishra, H. Sadia, A comprehensive analysis of fake news detection models: a systematic literature review and current challenges †, Eng Proc. (2023) https://doi.org/10.3390/engproc2023059028.

[32] S. Patel, S. Surati, MTL-rtFND: multimodal transfer learning for real-time fake news detection on social Media, in: Communications in Computer and Information Science, 2024.

[33] A. Al Obaid, H. Khotanlou, M. Mansoorizadeh, D. Zabihzadeh, Multimodal fake-news recognition using ensemble of deep learners, Entropy (2022), https://doi.org/10.3390/e24091242.

[34] R. Mohawesh, S. Maqsood, Q. Althebyan, Multilingual deep learning framework for fake news detection using capsule neural network, J. Intell. Inf. Syst. (2023), https://doi.org/10.1007/s10844-023-00788-y.

[35] R.K. Kaliyar, A. Goswami, P. Narang, FakeBERT: fake news detection in social media with a BERT-based deep learning approach, Multimed. Tools. Appl. (2021), https://doi.org/10.1007/s11042-020-10183-2.

[36] P. Dhiman, A. Kaur, D. Gupta, et al., GBERT: a hybrid deep learning model based on GPT-BERT for fake news detection, Heliyon 10 (2024).

[37] N.M. Duc Tuan, P. Quang Nhat Minh, Multimodal fusion with BERT and attention mechanism for fake news detection, in: Proceedings - 2021 RIVF International Conference on Computing and Communication Technologies, RIVF 2021, 2021.

[38] S. Singhal, R.R. Shah, T. Chakraborty, et al., SpotFake: a multi-modal framework for fake news detection, in: Proceedings - 2019 IEEE 5th International Conference on Multimedia Big Data, BigMM 2019, 2019.

[39] Y. Zhao, B. Hu, Y. Wang, et al., Identification of gastric cancer with convolutional neural networks: a systematic review, Multimed. Tools. Appl. 81 (2022) 11717–11736, https://doi.org/10.1007/s11042-022-12258-8.

[40] A. Giachanou, G. Zhang, P. Rosso, Multimodal multi-image fake news detection, in: Proceedings - 2020 IEEE 7th International Conference on Data Science and Advanced Analytics, DSAA 2020, 2020.

[41] J. Xue, Y. Wang, Y. Tian, et al., Detecting fake news by exploring the consistency of multimodal data, Inf. Process. Manage (2021), https://doi.org/10.1016/j.ipm.2021.102610.

[42] B. Palani, S. Elango, K. Vignesh Viswanathan, CB-Fake: a multimodal deep learning framework for automatic fake news detection using capsule neural network and BERT, Multimed. Tools. Appl. (2022), https://doi.org/10.1007/s11042-021-11782-3.

[43] H. Choi, Y. Ko, Effective fake news video detection using domain knowledge and multimodal data fusion on youtube, Pattern. Recognit. Lett. (2022), https://doi.org/10.1016/j.patrec.2022.01.007.

[44] P. Qi, Y. Bu, J. Cao, et al., FakeSV: a multimodal benchmark with rich social context for fake news detection on short video platforms, in: Proceedings of the 37th AAAI Conference on Artificial Intelligence, AAAI 2023, 2023.

[45] R. Kumari, V. Gupta, N. Ashok, et al., Emotion aided multi-task framework for video embedded misinformation detection, Multimed. Tools. Appl. (2024), https://doi.org/10.1007/s11042-023-17208-6.

[46] P. Palod, A. Patwari, S. Bahety, et al., Misleading metadata detection on youtube, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2019.

[47] H. Choi, Y. Ko, Using adversarial learning and biterm topic model for an effective fake news video detection system on heterogeneous topics and short texts, IEEe Access. (2021), https://doi.org/10.1109/ACCESS.2021.3122978.

[48] T. Mittal, U. Bhattacharya, R. Chandra, et al., Emotions don't lie: an audio-visual deepfake detection method using affective cues, in: MM 2020 - Proceedings of the 28th ACM International Conference on Multimedia, 2020.

[49] H. Khalid, M. Kim, S. Tariq, S.S. Woo, Evaluation of an audio-video multimodal deepfake dataset using Unimodal and multimodal detectors, in: ADGD 2021 - Proceedings of the 1st Workshop on Synthetic Multimedia - Audiovisual Deepfake Generation and Detection, co-located with ACM MM 2021, 2021.

[50] S. Knshnan, M. Chen, Identifying tweets with fake news, in: Proceedings - 2018 IEEE 19th International Conference on Information Reuse and Integration for Data Science, IRI 2018, 2018.

[51] S.R. Sahoo, B.B. Gupta, Multiple features based approach for automatic fake news detection on social networks using deep learning, Appl. Soft. Comput. (2021), https://doi.org/10.1016/j.asoc.2020.106983.

[52] K. Shu, X. Zhou, S. Wang, et al., The role of user profiles for fake news detection, in: Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2019, 2019.

[53] S. Raza, C. Ding, Fake news detection based on news content and social contexts: a transformer-based approach, Int. J. Data Sci. Anal. (2022), https://doi.org/10.1007/s41060-021-00302-z.

[54] R.K. Kaliyar, A. Goswami, P. Narang, S. Sinha, FNDNet – A deep convolutional neural network for fake news detection, Cogn. Syst. Res. (2020), https://doi.org/10.1016/j.cogsys.2019.12.005.

[55] P. Shah, Z. Kobti, Multimodal fake news detection using a cultural algorithm with situational and normative knowledge, in: 2020 IEEE Congress on Evolutionary Computation, CEC 2020 - Conference Proceedings, 2020.

[56] X. Cui, Y. Li, Fake news detection in social Media based on multi-modal multi-task learning, Int. J. Adv. Comput. Sci. Appl. (2022), https://doi.org/10.14569/IJACSA.2022.01307106.

[57] J. Wang, J. Zheng, S. Yao, et al., TLFND: a multimodal fusion model based on three-level feature matching distance for fake news detection, Entropy (2023), https://doi.org/10.3390/e25111533.

[58] R. Kumari, A. Ekbal, AMFB: attention based multimodal factorized bilinear pooling for multimodal fake news detection, Expert. Syst. Appl. (2021), https://doi.org/10.1016/j.eswa.2021.115412.

[59] Y. Wu, P. Zhan, Y. Zhang, et al., Multimodal fusion with co-attention networks for fake news detection, in: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, 2021.

[60] V. Kishore, M. Kumar, Enhanced multimodal fake news detection with optimal feature fusion and modified Bi-LSTM architecture, Cybern. Syst. (2023), https://doi.org/10.1080/01969722.2023.2175155.

[61] H. Yang, J. Zhang, L. Zhang, et al., MRAN: multimodal relationship-aware attention network for fake news detection, Comput. Stand. Interfaces. (2024), https://doi.org/10.1016/j.csi.2023.103822.

[62] S.M. Dwivedi, S.B. Wankhade, Deep learning based semantic model for multimodal fake news detection, Int. J. Intell. Eng. Syst. (2024), https://doi.org/10.22266/ijies2024.0229.55.

[63] L. Ying, H. Yu, J. Wang, et al., Multi-level Multi-modal cross-attention network for fake news detection, IEEe Access. (2021), https://doi.org/10.1109/ACCESS.2021.3114093.

[64] Q. Zhang, J. Liu, F. Zhang, et al., Hierarchical Semantic Enhancement Network for multimodal fake news detection, in: MM 2023 - Proceedings of the 31st ACM International Conference on Multimedia, 2023.

[65] Y. Wu, Y. Tang, C. Fan, Y. Liu, MDVT: a multi-modal fake News detection framework based on Vision Transformer, in: ACM International Conference Proceeding Series, 2023.

[66] H. Yang, J. Zhang, Z. Hu, et al., Multimodal relationship-aware attention network for fake news detection, in: Proceedings - 2023 International Conference on Data Security and Privacy Protection, DSPP 2023, 2023.

[67] H. Yang, X. Zhao, D. Sun, et al., Multi-modal fake news detection on Social Media with Dual Attention Fusion networks, in: Proceedings - IEEE Symposium on Computers and Communications, 2021.

[68] H. Zhou, Y. Liu, P. Xuan, Fake news detection based on pre-training and multi-modal fusion, Jisuanji Gongcheng./Comput. Eng. (2024), https://doi.org/10.19678/.issn.1000-3428.0066412.

[69] D. Khattar, M. Gupta, J.S. Goud, V. Varma, MvaE: multimodal variational autoencoder for fake news detection, in: The Web Conference 2019 - Proceedings of the World Wide Web Conference, WWW 2019, 2019.

[70] S.K. Uppada, P. Patel, B. Sivaselvan, An image and text-based multimodal model for detecting fake news in OSN's, J. Intell. Inf. Syst. (2023), https://doi.org/10.1007/s10844-022-00764-y.

[71] H. Choi, Y. Ko, Using topic modeling and adversarial neural networks for fake news video detection, in: International Conference on Information and Knowledge Management, Proceedings, 2021.

[72] M. Elpeltagy, A. Ismail, M.S. Zaki, K. Eldahshan, A novel smart deepfake video detection system, Int. J. Adv. Comput. Sci. Appl. (2023), https://doi.org/10.14569/IJACSA.2023.0140144.

[73] S.A. Shahzad, A. Hashmi, S. Khan, et al., Lip sync matters: a novel multimodal forgery detector, in: Proceedings of 2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA ASC 2022, 2022.

[74] K. Chugh, P. Gupta, A. Dhall, R. Subramanian, Not made for each other- audio-visual dissonance-based deepfake detection and localization, in: MM 2020 - Proceedings of the 28th ACM International Conference on Multimedia, 2020.

[75] Y. Zhou, S.N. Lim, Joint Audio-visual deepfake detection, in: Proceedings of the IEEE International Conference on Computer Vision, 2021.

[76] Y. Zhang, W. Lin, J. Xu, Joint audio-visual attention with contrastive learning for more general deepfake detection, ACM Trans. Multimed. Comput. Commun. Appl. (2024), https://doi.org/10.1145/3625100.

[77] W. Yang, X. Zhou, Z. Chen, et al., AVoiD-DF: audio-Visual Joint learning for detecting deepfake, IEEE Trans. Inf. Forens.Secur. (2023), https://doi.org/10.1109/TIFS.2023.3262148.

[78] M. Anas Raza, K. Mahmood Malik, Multimodaltrace: deepfake detection using audiovisual representation Learning, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2023.

[79] H. Ilyas, A. Javed, K.M. Malik, AVFakeNet: a unified end-to-end dense swin transformer deep learning model for audio–visual deepfakes detection, Appl. Soft. Comput. (2023), https://doi.org/10.1016/j.asoc.2023.110124.

[80] Y. Zhu, J. Gao, X. Zhou, AVForensics: audio-driven Deepfake video detection with masking strategy in self-supervision, in: ICMR 2023 - Proceedings of the 2023 ACM International Conference on Multimedia Retrieval, 2023.

[81] R.K. Kaliyar, A. Goswami, P. Narang, EchoFakeD: improving fake news detection in social media with an efficient deep neural network, Neural Comput. Appl. (2021), https://doi.org/10.1007/s00521-020-05611-1.

[82] Seghier MEA Ben, Truong TT, C. Feiler, D. Höche, A hybrid deep learning model for predicting atmospheric corrosion in steel energy structures under maritime conditions based on time-series data, Results Eng. (2025) 104417.

[83] Y. Pande, J. Chaki, Brain tumor detection across diverse MR images: an automated triple-module approach integrating reduced fused deep features and machine learning, Results Eng. 25 (2025) 103832.

[84] C. Raj, P. Meel, ConvNet frameworks for multi-modal fake news detection, Appl. Intell. (2021), https://doi.org/10.1007/s10489-021-02345-y.

[85] C. Song, N. Ning, Y. Zhang, B. Wu, A multimodal fake news detection model based on crossmodal attention residual and multichannel convolutional neural networks, Inf. Process. Manage. (2021), https://doi.org/10.1016/j.ipm.2020.102437.

[86] K.I. Roumeliotis, N.D. Tselikas, D.K. Nasiopoulos, Fake News detection and classification: a comparative study of convolutional neural networks, large language models, and natural language processing models, Futur. Internet 17 (2025).

[87] Y. Liang, T. Tohti, A. Hamdulla, Multimodal false information detection method based on Text-CNN and SE module, PLoS One (2022), https://doi.org/10.1371/journal.pone.0277463.

[88] Y. Guo, A mutual attention based multimodal fusion for fake news detection on social network, Appl. Intell. (2023), https://doi.org/10.1007/s10489-022-04266-w.
[89] P. Zhu, J. Hua, K. Tang, et al., Multimodal fake news detection through intra-modality feature aggregation and inter-modality semantic fusion, Complex Intell. Syst. (2024) 1–13.
[90] M. Jiang, C. Jing, L. Chen, et al., An application study on multimodal fake news detection based on Albert-ResNet50 Model, Multimed. Tools. Appl. (2024), https://doi.org/10.1007/s11042-023-15741-y.
[91] X. Zhang, T. Sun, C. Yang, et al., Rumor detection based on depth and breadth with tree-structured recursive neural networks, in: Proceedings of the International Joint Conference on Neural Networks, 2023.
[92] M. Gôlo, M. Caravanti, R. Rossi, et al., Learning textual representations from multiple modalities to detect fake news through one-class learning, in: ACM International Conference Proceeding Series, 2021.
[93] R. Jaiswal, U.P. Singh, K.P. Singh, Fake news detection using BERT-VGG19 multimodal variational autoencoder, in: 2021 IEEE 8th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering, UPCON 2021, 2021.
[94] Y. Gu, I. Castro, G. Tyson, Detecting multimodal fake news with gated variational AutoEncoder, in: Proceedings of the 16th ACM Web Science Conference, 2024, pp. 129–138.
[95] F. Shan, H. Sun, M. Wang, Multimodal social Media fake news detection based on similarity inference and adversarial networks, Comput. Mater. Contin. (2024) 79.
[96] G. Zhang, M. Gao, Q. Li, et al., Multi-modal generative DeepFake detection via visual-language pretraining with gate fusion for cognitive computation, Cognit. Comput. (2024) 1–14.
[97] J. Wang, S. Qian, J. Hu, R. Hong, Positive unlabeled fake news detection via multi-modal masked transformer network, IEEE Trans. Multimed. (2024), https://doi.org/10.1109/TMM.2023.3263552.
[98] B. Wang, Y. Feng, X.X. Cai, et al., Multi-modal transformer using two-level visual features for fake news detection, Appl. Intell. (2023), https://doi.org/10.1007/s10489-022-04055-5.
[99] P. Liu, W. Qian, D. Xu, et al., Multi-modal fake news detection via bridging the gap between modals, Entropy (2023), https://doi.org/10.3390/e25040614.
[100] F. Ghorbanpour, M. Ramezani, M.A. Fazli, H.R. Rabiee, FNR: a similarity and transformer-based approach to detect multi-modal fake news in social media, Soc. Netw. Anal. Min. (2023), https://doi.org/10.1007/s13278-023-01065-0.
[101] J. Wang, S. Qian, J. Hu, R. Hong, Comment-context dual collaborative masked transformer network for fake news detection, IEEE Trans. Multimed. (2023), https://doi.org/10.1109/TMM.2023.3330074.
[102] D.K. Sharma, B. Singh, S. Agarwal, et al., FakedBits- detecting fake information on social platforms using multi-modal features, KSII Trans. Internet. Inf. Syst. (2023), https://doi.org/10.3837/tiis.2023.01.004.
[103] V. Nair, J. Pareek, S. Bhatt, A knowledge-based deep learning approach for automatic fake news detection using BERT on Twitter, Procedia Comput. Sci. 235 (2024) 1870–1882.
[104] K. Yu, S. Jiao, Z. Ma, Fake news detection based on BERT multi-domain and multi-modal fusion network, Comput. Vis. Image Underst. (2025) 104301.
[105] S. Jindal, R. Sood, R. Singh, et al., NewsBag: a multimodal benchmark dataset for fake news detection, in: CEUR Workshop Proceedings, 2020.
[106] Z. Jin, J. Cao, H. Guo, et al., Multimodal fusion with recurrent neural networks for rumor detection on microblogs, in: MM 2017 - Proceedings of the 2017 ACM Multimedia Conference, 2017.
[107] E.F. Ayetiran, Ö. Özgöbek, An inter-modal attention-based deep learning framework using unified modality for multimodal fake news, hate speech and offensive language detection, Inf. Syst. 123 (2024) 102378.
[108] Lakzaei B., Chehreghani M.H., Bagheri A. (2025) A decision-based heterogenous graph attention network for multi-class fake news detection. arXiv Prepr arXiv250103290.
[109] Xu X., Yu P., Xu Z., Wang J. (2025) A hybrid attention framework for fake news detection with large language models. arXiv Prepr arXiv250111967.
[110] L. Hu, Z. Zhao, W. Qi, et al., Multimodal matching-aware co-attention networks with mutual knowledge distillation for fake news detection, Inf. Sci. (2024), https://doi.org/10.1016/j.ins.2024.120310.
[111] Y. Guo, H. Ge, J. Li, A two-branch multimodal fake news detection model based on multimodal bilinear pooling and attention mechanism, Front. Comput. Sci. (2023), https://doi.org/10.3389/fcomp.2023.1159063.
[112] H. Padalko, V. Chomko, D. Chumachenko, A novel approach to fake news classification using LSTM-based deep learning models, Front. Big. Data 6 (2024) 1320800.
[113] M. Luqman, M. Faheem, W.Y. Ramay, et al., Utilizing ensemble learning for detecting multi-modal fake news, IEEe Access. (2024), https://doi.org/10.1109/ACCESS.2024.3357661.
[114] S.A. Yousif, R. Jehad, T.Z. Abdulhameed, Preventing fraud: developing a hybrid deep learning model for rapid fake news detection, in: AIP Conference Proceedings, AIP Publishing, 2025.
[115] P. Bhardwaj, K. Yadav, H. Alsharif, R.A. Aboalela, GAN-based unsupervised learning approach to generate and detect fake news. Lecture Notes in Networks and Systems, 2023.
[116] P. Meel, D.K. Vishwakarma, HAN, image captioning, and forensics ensemble multimodal fake news detection, Inf. Sci. (2021), https://doi.org/10.1016/j.ins.2021.03.037.
[117] A. Malik, D.K. Behera, J. Hota, A.R. Swain, Ensemble graph neural networks for fake news detection using user engagement and text features, Results Eng. 24 (2024) 103081.
[118] M.P.S. Gôlo, M.C. de Souza, R.G. Rossi, et al., One-class learning for fake news detection through multimodal variational autoencoders, Eng. Appl. Artif. Intell. (2023), https://doi.org/10.1016/j.engappai.2023.106088.
[119] M.Y. Chen, Y.W. Lai, J.W. Lian, Using deep learning models to detect fake news about COVID-19, ACM Trans. Internet. Technol. (2023), https://doi.org/10.1145/3533431.
[120] M. Al-Yahya, H. Al-Khalifa, H. Al-Baity, et al., Arabic fake news detection: comparative study of neural networks and transformer-based approaches, Complexity (2021), https://doi.org/10.1155/2021/5516945.
[121] Q. Zhang, Z. Guo, Y. Zhu, et al., A deep learning-based fast fake news detection model for cyber-physical social services, Pattern. Recognit. Lett. (2023), https://doi.org/10.1016/j.patrec.2023.02.026.
[122] S. Deepak, B. Chitturi, Deep neural approach to fake-news identification, Procedia Comput. Sci. (2020).
[123] A.S. Mahdi, N.M. Shati, Utilizing graph neural networks for the detection of fake news through analysis of relationships among various social Media entities, in: International Conference on Innovations of Intelligent Informatics, Networking, and Cybersecurity, Springer, 2024, pp. 172–185.
[124] P. Yang, J. Ma, Y. Liu, M. Liu, Multi-modal transformer for fake news detection, Math. Biosci. Eng. (2023), https://doi.org/10.3934/mbe.2023657.
[125] J. Lv, X. Wang, C. Shao, TMIF: transformer-based multi-modal interactive fusion for automatic rumor detection, Multimed. Syst. (2023), https://doi.org/10.1007/s00530-022-00916-8.
[126] Y. Zhou, Y. Yang, Q. Ying, et al., Multi-modal fake news detection on social Media via Multi-grained information fusion, in: ICMR 2023 - Proceedings of the 2023 ACM International Conference on Multimedia Retrieval, 2023.
[127] S. Kalra, Y. Sharma, P. Vyas, G.S. Chauhan, FakeRevealer: a multimodal framework for revealing the falsity of online tweets using transformer-based architectures, in: International Conference on Pattern Recognition Applications and Methods, 2023.
[128] W. Li, F. Bu, Multi-modal semantic enhancement for early fake news detection, in: 2023 4th International Conference on Electronic Communication and Artificial Intelligence, ICECAI 2023, 2023.
[129] P. Singh, R. Srivastava, K.P.S. Rana, V. Kumar, SEMI-FND: stacked ensemble based multimodal inferencing framework for faster fake news detection, Expert. Syst. Appl. (2023), https://doi.org/10.1016/j.eswa.2022.119302.
[130] M.I. Nadeem, K. Ahmed, D. Li, et al., EFND: a semantic, visual, and socially augmented deep framework for extreme fake news detection, Sustainability (2023), https://doi.org/10.3390/su15010133.
[131] A. Kishwar, A. Zafar, Fake news detection on Pakistani news using machine learning and deep learning, Expert. Syst. Appl. (2023), https://doi.org/10.1016/j.eswa.2022.118558.
[132] E.D. Ajik, G.N. Obunadike, F.O. Echobu, Fake news detection using optimized CNN and LSTM techniques, J. Inf. Syst. Inform. (2023), https://doi.org/10.51519/journalisi.v5i3.548.
[133] F. Folino, G. Folino, M. Guarascio, et al., Towards data-and compute-efficient fake-news detection: an approach combining active learning and pre-trained language models, SN Comput. Sci. 5 (2024) 470.