SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Advances And Challenges In Multimodal Deep Learning For Social Media Fraud And Misinformation Detection

Mr. Omganapathy A[1], Dr. V J Chakravarthy[2], Dr. France K[3]

[1]Research Scholar Faculty of Computer Applications Dr.M.G.R.Educational and Research Institute[1] Chennai, Tamil Nadu, India.
omganapathy.mca@drmgrdu.ac.in
[2]Professor Faculty of Computer Applications Dr.M.G.R.Educational and Research Institute Chennai, Tamil Nadu, India.
chakravarthy.mca@drmgrdu.ac.in
[3]Assistant Professor Department of Computer Science and Engineering Aarupadai Veedu Institute of Technology Vinayaka Mission Research Foundation Deemed University Chennai – 603104
franceapm@gmail.com

Abstract
The rapid growth of social media has paved the way for the proliferation of deceptive and misleading content, thereby making it essential to develop trustworthy and automated ways. This systematic review is concerned with the evolution of the use of multimodal deep learning methods. For detecting fraud and misinformation on social media, highlighting aspects like datasets modalities ways, performance level, research gaps, and emerging trends. Articles which contain several types of content like text images audio, videos, social media features, and behavioural patterns. Also present a mix of machine learning techniques, deep learning, to advanced models like multimodal fusion attention mechanisms. The evaluation done here has revealed that usually there has been an evolution from traditional single-modal or text-based machine learning to multimodal, contextual, and deep architectures. The social media datasets most extensively utilized in this domain are Twitter, Weibo, and fakeddit. Most multimodal combinations still involve either text-image or text-social-context pairings. Another outcome is the realization that multimodal integration is still quite fragmentary, as is dataset, evaluation, and even multimodal fusion coverage. The survey has pinpointed a research gap that the lack of unified systems that, at the same time, model the content user behaviour social links, external knowledge, and the time context. Because of this, the paper offers a setup for a systematic knowledge of the state of the art and proposes ways forward aimed at developing more holistic, scalable, and trustworthy multimodal detection systems.

Keywords: Social media, Datasets modalities, Machine learning techniques, Fraud and misinformation, Deep learning, Multimodal fusion

1. Introduction

(i) Evolution of Multimodal Misinformation Detection

Multimodal approaches to fraud and misinformation detection have been able to achieve considerable results through the fusion of textual, visual, and contextual cues obtained from different social media platforms [1]. The increasing spread of fake news in social media and the employment of AI techniques for the detection of false news in the language and the different modalities. They show that the previous reviews have missed some aspects like covering recent trends datasets representation learning, and classification methods [2]. Social media growth has been at an extremely fast rate. The result unfortunately is the proliferation of fake news, misinformation, and fraudulent content across different platforms. With the increased use of multimedia content - including but not limited to text images audio and video accurate identification of misleading content is becoming one of the biggest challenges [3]. With the fast rise of fake news on social media spreading so quickly, there has been an increased demand over time for dependable identification techniques that can easily separate true facts from false ones. In the field of machine learning, deep learning, and multimodal analysis, these cuttingedge fields are the ones that have allowed the combination of textual, visual, and contextual features to come up with the best results when it comes to fake news detection [4].

(ii) Multimodal Learning for Fraudulent Activity Detection

Fake accounts play a substantial role in spreading misinformation, spam, and malicious actions over social media platforms. Detection with the help of automated tools is the only feasible option here due to the difficulty of manually identifying fake accounts mainly at a highly sophisticated level [5]. Fake news detection (FND) aims to assess the credibility of information and identify false or misleading content. The rapid spread of misinformation across social media in textual and visual forms has increased the need for accurate and reliable detection methods. Therefore, multimodal deep learning approaches are gaining importance for effectively analyzing diverse content and improving misinformation detection [6]. Fake news detection methods can be broadly classified into two categories unimodal and multimodal approaches**.** Unimodal methods analyze a single type of information, such as text, images, or videos, whereas multimodal methods combine multiple sources, including text, images, and social-contextual information [7]. Traditional single-channel methods usually fail to model the intricate interactions of different forms of media content with user behavior on social platforms. Multimodal deep learning, which takes into account not only text but as well images, audio, and social factors, gives a more complete view that enhances detection of frauds and fake news [8].

(iii) Emerging Threats and Challenges in Social Media Fraud Detection

The rising sophistication of activities on social media has rendered spotting fraud and coordinated behavior almost impossible for traditional means. Combining the use of different modalities, deep learning methods can exploit various forms of information like temporal textual visual, audio, and behavioral data to find new types of patterns and manipulations with multiple coordinated actors [9]. Multimodal deep learning offers an alternative by incorporating data from various types of data like transactional, behavioral, and contextual levels into one comprehensive system. This fusion of different modalities in detection methods not only increases detection accuracy, but also helps system adaptability and transparency [10]. Digital platforms are used more and more, which has enabled the fraudsters who are making increasingly complicated fraud, such as fake faces of video, video of fake facial expressions, voice alterations, and fake coordinated human behavior patterns [11]. Phishing attacks continue to evolve, becoming increasingly sophisticated and difficult to detect using traditional single-modal and static analysis techniques. Multimodal deep learning can address these limitations by integrating textual, visual, behavioral, and temporal information to capture complex patterns associated with fraudulent activities. Such approaches provide a promising direction for developing more accurate and adaptive fraud detection systems for social media environments [12].

1.1. Contribution of review

This review paper highlights a detailed overview of multimodal techniques used for fraud and misinformation detection in social media platforms. The primary contributions of this paper are:

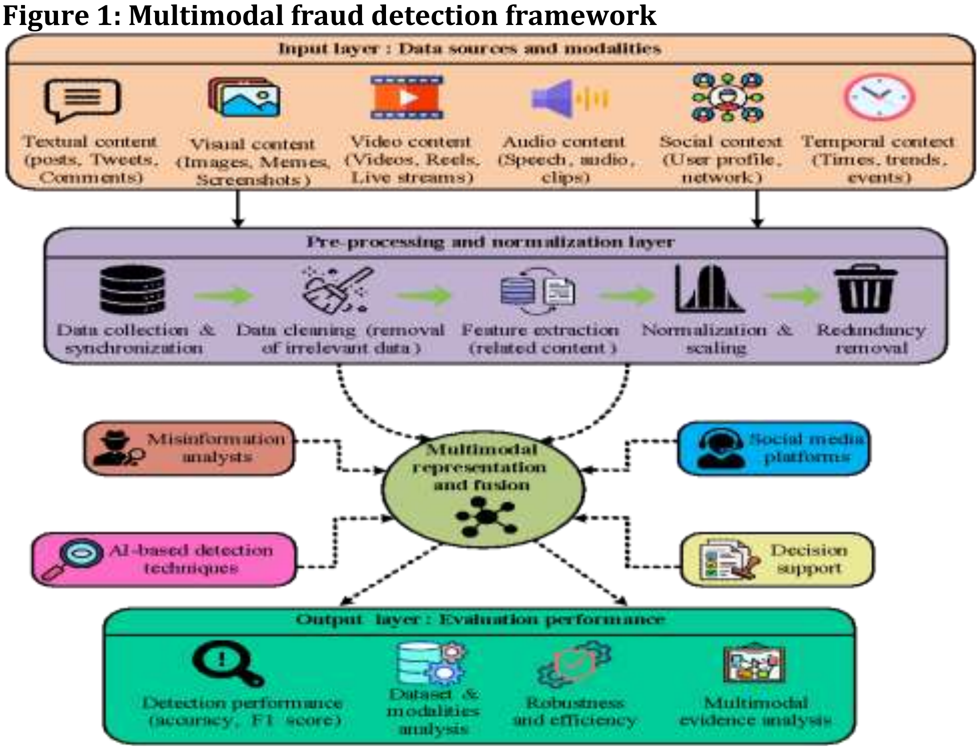
- **Comprehensive Multimodal Analysis:** This survey discusses the recent approaches to deep learning which combine the use of text images audio, video, and other contextual features.
- **Taxonomy of Methods and Datasets:** The survey classifies the existing studies based on datasets modalities representation learning, fusion strategies, and detection methods.
- **Research Gaps and Challenges:** Mainly the problems such as data scarcity, cross-modal inconsistency explanation generalization, deep fakes, and on-line detection are identified.
- **Future Research Directions:** New trends in research like multimodal learning via a large language model (LLM)-based structure, adaptive fusion techniques, explainable artificial intelligence (AI), detection across platforms, and development of robust real-time systems.

The rest of the paper is arranged in this manner: Section 2 elaborates the framework of multimodal detection in fraud and misinformation. The overview of the systematic literature is presented in Section 3. It also includes the comparative analysis of existing datasets, modalities, and detection methods. Section 4 explains about the research evolution and taxonomy of the fraud and misinformation detection. Section 5 highlights the result analysis of the fraud and misinformation detection. Section 6 talks about the research gap and the challenges faced in the reviewed studies. Finally, in Section 7 concludes the review and elaborates the future research directions of the findings.

2. Overall framework of multimodal detection in fraud and misinformation

The figure 1 shows the overall framework for the detection of multimodal fraud and misinformation in social media. At the input layer, the heterogenous data sources, including textual pictorial video audio social-context, and temporal-context information, are combined and integrated. Data undergoes pre-processing and the normalization step that consist of collecting and aligning data, removing irrelevant data, extracting related content, scaling features, and reducing redundancies. The modalities that were processed are transformed into identical representations with the help of multimodal representation and fusion so that the complementary cross-modal features could be integrated for the better detection. The system gives support to identifying and analyzing potentially fraudulent/ misleading contents. Eventually, the layer of outputs judges the different reviewed methods based on the factors of detection performance, data, and modality description robustness and computational efficiency, multimodal evidence analysis so that the effectiveness of and research trends in multimodal detection can be made evident from a comprehensive overview.

Figure 1: Multimodal fraud detection framework



3. Systematic literature review

The Systematic Literature Review is a way for finding, analyzing and synthesizing existing literature on a selected topic. It can be used to analyze the existing literature on multimodal deep learning for fraud and misinformation detection on social media, the datasets used modalities models, fusion methods, evaluation scores used etc. It can also help identify gaps in existing research and areas of trends and limitations and potential future research.

3.1. PRISMA-Based Systematic Literature Review Framework

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guide in this paper to describe, in sequence, how potential studies that investigate multimodal deep learning-based fraud and misinformation detection on social media platforms are identified and selected. This setup outlines how to systematically and transparently carry out these activities so that readers can follow the steps taken.

The research literature review comprised searching the main scientific databases of IEEE Explorer, ScienceDirect, SpringerLink, ACM Digital Library, Web of Science, and Google Scholar. A period of 2022-2026 was the timeframe used to capture a recent status of developments on a topic of multimodal deep learning as a countermeasure for fraud and misinformation on social networks.

For the initial search keywords, different groupings were used with Boolean operators

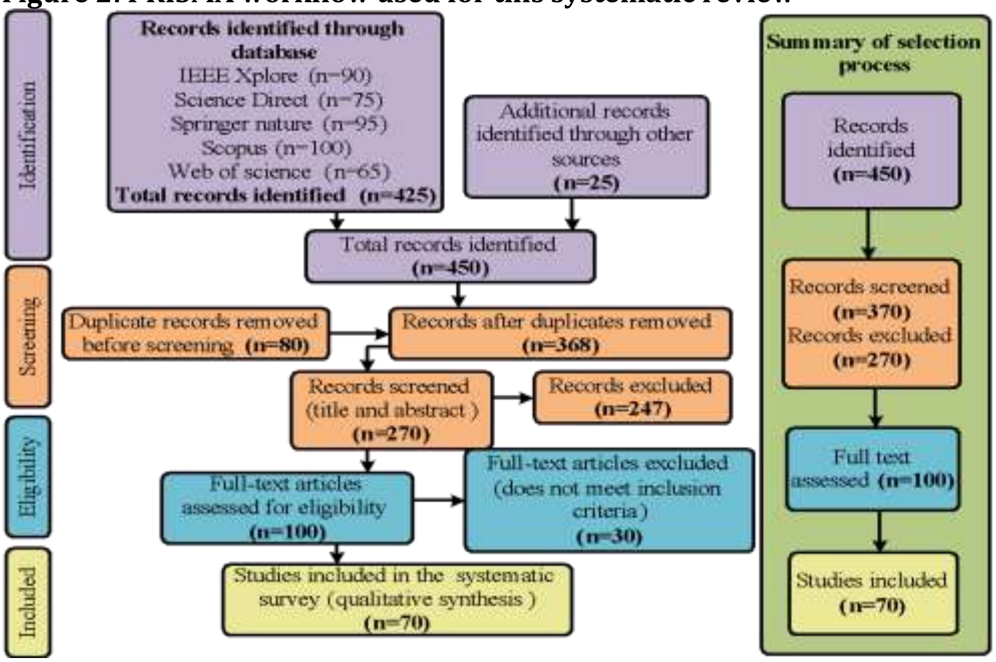
Search String 1:
("Multimodal deep learning "AND" Misinformation detection")
Search String 2:
("Fake news detection" AND "Social media")
Search String 3:
("Multimodal fraud detection" AND "Social media")
Search String 4:
("Deep learning" AND "Fraud detection").

Table 1: Exclusion and Inclusion Used in PRISMA-Based Literature Selection.

Criteria	Inclusion Criteria	Exclusion Criteria
Type of Publication	Journal articles and review papers	Conference papers, editorial books, Book chapters, theses
Language	English	Publications in other languages
Publication Year	Publications within 2022-2026	Publications before 2022
Research Area	Multimodal deep learning, fraud detection, misinformation detection, fake news, social media	Unrelated domains and non-social-media applications
Data Modality	Text image video, audio, and social/behavioral information	Studies using unrelated or insufficient data modalities
Way	Machine learning, deep learning, multimodal fusion transformers LLMs, and related AI techniques	Non-AI or purely rule-based approaches
Content	Studies addressing fraud misinformation fake news, fake accounts, or related social-media threats	Duplicate studies, irrelevant studies, and studies with insufficient information

In table 1, the retrieved articles were reviewed and evaluated using the preset criteria of inclusion and exclusion. The PRISMA guideline was followed to guarantee a rigorous and transparent selection of papers. It is carried out a review with four phases: Identification Screening Eligibility, and Inclusion. After paper selection categorized them based on datasets, data modalities, deep learning architectures, fusion strategies, evaluation metrics, and research trends limitations future research directions, etc. Figure 2 elaborates the PRISMA flow diagram is basically explaining a stepwise selection of studies that were used in the review. In the first phase, 450 records were retrieved from the relevant databases. Then, 370 unique records underwent title and abstract screenings after removing the duplicates. From these matches, 270 records were removed for not being relevant to the concepts of social media misinformation or fraud based on the inclusion/exclusion criteria in place. Following this, 100 articles were selected for the review. 30 out of those were excluded for heavily relying on unimodal models, absence of baseline performance indications, or absence of behavioral/contextual traits.

Figure 2: PRISMA workflow used for this systematic review



Moreover, the Research Questions (RQ) are mentioned as below,

RQ1: Which multimodal deep learning methods and models have been designed for identifying fraudulent information or fake news in social networking contexts?

RQ2: What kind of data collection fusion analysis and other aspects are typically involved when multimodal methods are applied to detect fraudulent activities and misinformation from social media?

RQ3: What are the primary difficulties, research needs, new directions, and prospective studies for the identification of fraud and misinformation through multimodal deep learning techniques on social media?

The above mentioned RQs are carefully analyses for their process and relevant performance based on the goals. Also, research finding and objectives are reviewed with respect to the database in terms of keywords. Furthermore, the Research Objectives (RO) is summarized as follows;

RO 1: To systematically investigate various multimodal deep learning methods and models that have been developed to detect fraud and misinformation on social media platforms.

RO 2: To conduct an assessment of the datasets, modalities of data, ways of feature extraction, strategies for fusion, and criteria for evaluation of the performance that have been used in previous works.

RO 3: To highlight main areas of research that have not been sufficiently addressed yet, to outline trends that seem to have recently emerged in the field, to discuss the challenges and to point out the directions of research that could be done to build multimodal fraud and misinformation detection systems with high reliability.

3.2. Answer for RQ1: various multimodal deep learning methods and models that have been developed to detect fraud and misinformation on social media platforms

Based on findings from these studies obviously multimodal deep learning has been recognized as a key method in social media fraud and misinformation detection that is achieved by joint analysis of text image sound and user behaviors.

Aditya and Mohanty [13] created a heterogeneous Word2Vec with CNNs to detect fake-profiles, incorporating the different modalities of text image video, audio, and behavioral data. And then the GWO and EHO are used to extract the features. Che et al., [14] introduced Sparse and Graph-Regularized CANDECOMP /PARAFAC (SGCP) method to automatically identify news items that users have shared actually fake by leveraging information about the social network. This method employs a sparse and graph regularization combined with monotonic optimization to enhance fake-news detection accuracy.

Sharma and Garg [15] introducing the Indian Fake News Dataset (IFND) that consists of textual and visual news content. LDA was applied to topic modeling intelligent augmentation was also employed to generate more fake-news examples. Uppada et al., [16] presented Social Engagement-based News Authenticity Detection (SENAD) as the new method to detect fake-words or fake- accounts the feature of user. Developed a Credibility Neural Network (CredNN) as well, which used a CNN-based image classification, Error Level Analysis (ELA), and sentiment analysis to catch fake or manipulated images. Park and Chai [17] introduced a machine-learning setup centered on the user, combining features from the user, news content, and social networks for fakeNews detection. XGBoost was employed for feature-importance analysis then SVM, Random Forest (RF), Logistic Regression (LR), CART, and NNET classifiers.

Altheneyan and Alhadlaq [18] presented a distributed stacked-ensemble-based machine-learning structure utilizing Apache Spark. It is used in scalable fake-news detection. This approach first obtains the features from three methods: N-grams, Hashing TF-IDF, and Count Vectorizer. Guo et al., [19] have introduced a Multi scale Transformer-based fake-news detection technique. Uppada et al., [20] presented an image, text multimodal scheme for spotting fake-news on social media that is also based on the Fakeddit dataset by utilizing a combination of visual and linguistic aspects from posts. Zhou et al., [21] propose CLIP-GCN, a method for fake news detection that fuses multimodal data. The model is built using CLIP for extracting fused semantic image-text features, GCN for encoding intra-domain knowledge. Gedara et al., [22] presented an innovative multimodal fake-news detection system combines Fuzzy C-Means, AutoGluon, and LLM-based prompting for image and text processing at the same time. Mbaziira et al., [23] highlighted the study that has been the use of multimodal AI/ML and light detection methods as a way of mitigating multimedia misinformation, cross-lingual content issues, and real-time social-media dynamics.

Nikumbh and Thakare [24] evaluated the multimodal fake news detection model which fuses BERT-LSTM text features with CNN image editing features and MLP social contextual features. Mouratidis et al., [25] elaborates fake-news detection using various classifiers like Nave Bayes SVM Random Forest CNN LSTM, and BERT by training the features provided by different word-representation schemes such as TF-IDF, Word2Vec, and contextual. Seddari et al., [26] proposed a multimodal deep-learning setup that jointly leveraged text and image modalities for misinformation detection in multimedia posts in an end-to-end way. Al-Alshaqi et al., [27] introduced a BERT- based multimodal fake news detector model that brings together original text content with OCR-extracted text from images.

Raza et al., [28] demonstrated a bibliometric analysis of disinformation detection research that uses machine-learning and deep-learning. Web of Science, Biblioshiny 4.2.0, n-gram analysis, and mixed analysis, were utilized by the study to recognize trends, important works, and new directions. Zhao et al., [29] introduced DReMHIM, a multimodal deep reinforcement learning model that joins RoBERTa for text, ResNet-50 for pictures, and dual-attention fusion to spotting health misinformation. Abdullah et al., [30] presented a multimodal ensemble structure based on ViLBERT to generate visual-textual representations. The features obtained are categorized into different classes by several models like Bidirectional Gated Recurrent Unit (Bi-GRU), Transformer-XL, DistilBERT and XLNet. The table 2 summarizes the multimodal deep learning techniques for fake news detection in social media.

Table 2: Summary of Multimodal Deep Learning Techniques for Fake News Detection in Social Media

Reference	Integrated techniques	Dataset	Strength	Weakness	Outcome
Aditya and Mohanty [13]	bCNN Cosine Convolutional, Gabor & Wavelet transforms GWO EHO, 1D-CNN Word2Vec	Fourier Diversified internet data (text photo audio/video, social activity)	Analyzing fake profiles from heterogeneous sources	Computationally complex-focuses on fake user profile verification instead of fake news identification	Accurate detection of fake profiles, reported an accuracy of 93.5% for detecting fake login/signup
Che et al., [14]	Multimodal sparse-regularized model, social regularization, tensor decomposition	News consumption on social media	Drawing upon both news content and the underlying social network to get a multi-modal understanding of the problem	Multimedia resources in the form of video and images are not directly used	Utilizing social context info as an effective way to spot fake stories
Sharma and Garg [15]	Integration of text & image ML/DL classifiers, intelligent augmentation	Multimodal fake news dataset from India Indian Fake News Detection Dataset(IFND)	Generalization of performance can be severely influenced by the quality and composition of the dataset	Generalization of performance can be severely influenced by the quality and composition of the dataset	The work allows testing of fake news models based on text, visuals and their combination

Uppada et al., [16]	Twitter/OSN social data, image data, features based on social interaction, authenticity score, features of social engagement;	CredNN CNN ELA , sentiment analysis combination of user behavior and visual features	Twitter/OSN social data, image data, features based on social interaction, authenticity score, features of social engagement;	Detection methods are separated in different tasks	FALSE SENAD-93.7% accuracy, nCredNN-approximately 76% accuracy
Park and Chai [17]	XGBoost SVM RF LR CART, NNET;	User, news-content and social-network features	User-centered social-capital-based detection	Conventional ML has weakness in deep semantic/ multimodal representation of social media content	RF with an accuracy of about 94% is obtained
Altheneyan and Alhadlaq [18]	Apache Spark, N-grams, Hashing TF-IDF, Count Vectorizer, stacked ensemble	FNC-1	Distributed and scalable fake-news detection	Mostly text-based; need distributed computing	92.45% F1-score
Guo et al., [19]	Multiscale Transformer	Mixed-language social-media text	During mixed-language text classification process, understanding multi-level semantics information	Text only; visual modality information are ignored	roughly 2-10% more accurate than baseline models
Uppada et al.,[20]	Two distinct textual and visual models with multimodal feature fusion	Fakeddit, Image polarity dataset taken from Flickr	Capitalizes on the context of both linguistic and visual information together	Mainly constrained to text-image modality	91.94% Accuracy, 93.43% Precision, 93.07% Recall, 93% F1
Zhou et al., [21]	CLIP, Adversarial Neural Network, GCN	Weibo, Twitter	Comprising multimodal emergent web news with fewer labels	Domain adaptation and graph construction adds to the complexity	88.7% average accuracy

Gedara et al., [22]	Fuzzy C-Means AutoGluon LLM prompt learning, Functional Network	Three benchmark multimodal datasets	Multimodal detection with explainability	Computationally expensive and complicated process	10% average accuracy increase
Mbaziira et al., [23]	Multimodal AI/ML, multilingual & light Weight detection methods	Multimodal dataset	Focused on multimodality, multilingual and real-time detection	Fails to systematically define a single proposded architecture	outlines requirements and future direction for efficient detection
Nikumbh and Thakare [24]	BERT-LSTM CNN MLP, scaled dot-product attention, feature fusion Textual visual and social-context information	Fake News Net, Fake edit ,Twitter, Media Eval	Exploit textual, social and visual contexts information	Complex multimodal architecture	Multimodal fusion performance outperformed the detection of unimodal approaches.
Mouratidis et al. [25]	Nave Bayes SVM RF CNN LSTM, BERT; TF-IDF Word2Vec contextual embeddings	Several fake-news datasets	Extensive comparison of traditional and DL models	Not a new multi-modal architecture; mostly comparison	Overall BERT models performed the best
Seddari et al [26]	Multimedia deep learning model; Text-Image features extract-and-fuse	Multimedia social-media content	The model combines text and image modalities	Not able to derive precise architecture from given information	Illustrates the crucial role of multimodal input for uncovering fake news
Al-Alshaqi et al., [27]	BERT base uncased + OCR, OCR-extracted image and text	ISOT WELFAKE TRUTHSEEKER, and a composite dataset	The model combines written text and images with the help of OCR, which extracts text from images	OCR can only extract the textual part of images	99.97% accuracy, 0.981 F1 score

Raza et al., [28]	Biblioshiny 4.2, n-gram analysis and hybrid analysis	Web of Science	A model that detects the evolution of topics through identification of research trends, the influence of papers, or the emergence of new fields,	No new detection or an experimental classifier	Announced research in misinformation via ML/DL increased by 96.1% per year
Zhao et al., [29]	RoBERTa, ResNet-50, a pair of attentions, and a Rainbow DQN	Over 3 billion tweets plus 31,282 Instagram captions	Multimodal learning and deep reinforcement learning are combined here, with an emphasis on credibility	Extremely data-intensive and computation-demanding	Improvement in detection by 12.5%, while spread of false information reduction by 45.2%
Abdullah et al., [30]	ViLBERT VADER Image - Text Contextual Similarity, Bi-GRU, Transformer-XL DistilBERT XLNet stacked soft voting, T5	Fake News and Social Commentary Dataset, Weibo	An innovative multimodal integration with cross modal inconsistency as the core detection component	The model is very heavy on both computation and memory	At most 96% accuracy on Fakeddit, and over 94% on Weibo were reported.

3.3. Answer for RQ 2: Types of datasets collected to detect fraudulent activities and misinformation from social media

According to the analyzed publications text images videos audio, as well as social/behavioral information are the primary modalities used for fraud and misinformation detection on social media. Francis et al. [31] developed a multimodal fake detection Tamil model which used 7,934 fact-checked samples that were mixed with text, images, and speech. They used features specific to each modality through BERT, ViT/ResNet and MFCC/HuBERT. Bhukya et al. [32] developed multiple media modalities and is trained on Twitter, Weibo, and Fakeddit data. Features are extracted by a pretrained model, which is RoBERTa, from the text and a pretrained ViT model for the images. Zamil and Charkari [33] used the Weibo, MediaEval, and CASIA datasets. To get the text features, used ELECTRA/XLNet.

Hajli et al. [34] combined LLM and machine learning solutions that proposes cross-modal analysis and adaptive learning. Abdulridha and Albaker [35] detect real-time multimodal scenario by using videos and audios. Features like those related to the visual, acoustic, and speech was first extracted and combined using techniques of machine learning. Hamed et al. [36] introduced DA-LGS, a GAN-based data augmentation technique, which they used on Fakeddit dataset to detect fake news with the use of a fine-grained multimodal method. Abbas and Taeihagh [37] proposed a multi-level fusion structure via CNN and RNN for retrieving deep visual and semantic textual features together for multimodal fake-news classification. The structure achieved up to 98.16% accuracy on the WELFake data set. Wu et al. [38] introduced the multimodal dual-CNN named MDCNN, which combines audio and text for emotion features to identify fake news. The IEMOCAP dataset combine text for multimodal fake news identification. LekshmiAmmal and Madasamy [39] developed an interpretable Tamil fake-news detection model utilizing multimodal cues, as both text and visuals were obtained from collaborative fact checking and official news platform. Joshi et al. [40] introduced an interpretable cross-platform misinformation detection structure with DANN to learn domain-invariant features of multiple social-media platforms.

Wang et al. [41] have been able to build in SEPM an incremental multimodal fusion networks for counterfeit information detection with textual and visual content from twitter and weibo datasets. Fu and Liu [42] evaluated dataset Twitter15 and Twitter16 to detect the fake news with Twitter15 90.5% accuracy and Twitter16 with 93.0%. Aishwarya et al. [43] introduced a BiLSTM with attention mechanism setup for classifying text-based misinformation. Shukla et al. [44] proposed an account detection system using TCN, data augmentation via GAN, and Autoencoder based feature extraction. The system was evaluated on a significant dataset Cresci-2017and TwiBot-22 and ROC-AUC scores that showed good performance to analyze sequential user-behavior data. Kozik et al. [45] evaluated a six benchmark datasets demonstrated promising results of the deep-learning system. Table 3 elaborates the types of datasets collected to detect fraudulent activities and misinformation from social media.

Table 3: Comparative Analysis of Datasets, and Detection Performance for Social Media Fraud and Misinformation Detection

Author	Types	Integrated models	Reported Detection Performance	Datasets used
Francis et al.[31]	Text + Image + Speech	BERT/IndicBERT+ ViT/ResNet/VGG16 +MFCC /HuBERT + cost sensitive learning	Accuracy 76.08%, F1 0.85	TamilFacts 7,934 fact checked samples
Bhukya et al. [32]	Text + Image + Behavioural	RoBERTa + ViT + reliability-aware adaptive late fusion + MLP	Twitter 99.69%, Weibo 98.93%, Fakeddit 85.14%	Twitter, Weibo , Fakeddit
Zamil and Charkari [33]	Text + Image	ELECTRA + XLNet + ELA + EfficientNet B0 + LIME	Accuracy 82.08%	Weibo,, MediaEval, CASIA
Hajli et al. [34]	Text + Propagation + Multimodal	LLMs (BERT, GPT, LLaMA) + ML + cross-modal analysis	Accuracy 78.67%, F1 0.94	Multiple datasets discussed
Abdulridha and Albaker [35]	Video + Audio + Linguistic	Visual + acoustic + linguistic feature extraction + ML + parallel computing	99.83% accuracy	Multimodal deception datasets / experimental scenarios
Hamed et al.[36]	Text + Image	LeakGAN + GAN + StackGAN + DA-LGS	Accuracy 90.10%	Fakeddit
Abbas and Taeihagh [37]	Text + Image	Dual-CNN + RNN + mean/weighted/max/sum fusion + polynomial-kernel classifier	WELFake 98.16%	ISOT, Fake vs Real news WELFake FA-KES, Twitter

Wu et al [38],	Audio + Text	Dual-CNN + recursive encoder + feature fusion + speech-emotion analysis	WAP 78.8%	IEMOCAP
LekshmiAmmal and Madasamy [39],	Text Image	Transformers + LLM image captions + Siamese network + XAI	F1- 0.8736	Tamil fact-checking and mainstream news organizations
Joshi et al. [40]	Text/social-domain information	DANN + LIME	Accuracy +3%, AUC +9%	Two COVID-19 misinformation datasets
Wang et al. [41]	Text + image	SEPM +semantic enhancement + progressive/hierarchical multimodal fusion	Improvement 3.5% Twitter, 2% Weibo	Twitter, weibo
Fu and Liu [42]	Text + User interaction + Knowledge	Knowledge Graph + weighted GCN (A-KWGCN)	Twitter15 - 90.5%, Twitter16-93.0%	Twitter, weibo
Aishwarya et al. [43]	Text	BiLSTM + Attention	Accuracy 87.97%	Benchmark fake/real news datasets
Shukla et al. [44]	User behavior, Account activity	TCN + GAN + Autoencoder + SOA	ROC-AUC 96 / 95	Cresci-2017 - TwiBot-22
Kozik et al. [45]	Text	Deep-learning classifier committee / ensemble learning	Accuracy 94.07%	Six benchmark datasets

3.4. Answer for RQ 3: Challenges, Research Gaps, Emerging Trends, and Future Directions in Multimodal Deep Learning for Social Media Fraud and Misinformation Detection

According to the literature reviewed, the key challenges in detecting multimodal fraud and misinformation revolve around integrating diverse and heterogeneous data sets, aligning across different modalities, scarcity of data, imbalanced datasets, constantly changing misinformation techniques, adversarial hacking impersonators privacy-related issues, and computationally expensive demands of the models. Guo and song [46] mentioned several major challenges in social media analytics like data heterogeneity, scalability, computational cost, and real-time processing. Also, the study pointed out GNNs and self-supervised learning as possible areas. Alhakami et al. [47] have discussed about the algorithmic bias, echo chambers, scalability, cross-platform adaptability, and computational efficiency as main issues hampering AI-based social-media analysis. Xia et al. [48] suggested a new method for disinformation detection in social media the explainable BERT-LSTM setup. These studies demonstrate the importance of transparent, interpretable, and feature-aware models in the fight against misinformation. Sudhakar and Kaliyamurthie [49] carried out a study to compare different techniques of identifying misinformation related to the spread of Covid-19, namely "machine-learning and deep-learning models with a gigantic dataset of "1,375,592 tweets. Jarrahi and Safari [50] pointed out that, when the main task is to detect fake news on social media, the characteristics such as those of the publisher and user are crucial. Using their FR-Detect system, they combined a publisher's attributes like credibility influence sociality, validity and the length of the existence and also textual information to achieve the goal with a sentence-level CNN.

Singh et al. [51] drew attention to the increasing challenge of distinguishing between AI-generated and human-generated disinformation, and called for the development of strong, adaptable detection models. Their model combining different techniques and BERT demonstrated very good results. They emphasized that future efforts should focus on the detection of synthetic-content, real-time checking of facts, scalability, and ability. Goyal et al. [52] uses LSTM for the detection of fakes through text, CNN for fake detection through visual content, and ANN for the utilization of account metadata. This work is still relevant to the area of research related to large scale multimodal learning, fake accounts behavior and graph analysis, and adaptive detection of fake accounts which are highly sophisticated. Tian et al. [53] pointed out the importance of achieving equilibrium among these components of automated fake-news detection accuracy, interpretability, computational efficiency, and generalization. And also highlighting the importance of further research is efficient, interpretable, and context-aware. Sharma et al. [54] also listed major problems like shortage of high-quality datasets, not knowing well how users behave, limitations of natural language processing tools, and emotionally manipulating content as major constraints. Sittar et al. [55] tackled the problem of data imbalance and multilingual, cross-media issues through LLM-driven text augmentation.

Xu and kechadi [56] highlight the need for models that are uncertainty-aware, better benchmark datasets, and thorough evaluation. It is recommended that future research work on ambiguous texts and improving the reliability of models for misinformation detection in real-world scenarios. Li et al. [57] emphasized that in their work LLMs have both positive and negative roles in detecting misinformation, since they can be useful for fact checking but at the same time capable of creating very convincing fake information. In particular, the paper called for further investigation in trustworthy LLMs, stronger fact-checking mechanisms, ethical deployment and regulatory oversight. Premkumar and Nachimuthu [58] put forward an innovative privacy-preserving federated cross-modal graph transformer to combine decentralized social-media threat detection through textual visual audio and social-graph input. The researchers identified some of the areas for improvement such as privacy, scalability, and multimodal features. Sormeily et al [59] pointed out issues with limited early propagation data and also discussed features like users, temporal changes, and spreading patterns. It emphasizes the coming necessity for a kind of early-stage detectors that are aware of news spreading, handle multiple channels, and operate in real-time. Nadeem et al. [60] emphasize the importance of semantic analysis across different modalities and efficient detection of forgery to increase robustness of the system in complex scenarios of social networking. The table 4 summarizes the Challenges, Research Gaps, Emerging Trends, and Future Directions in Multimodal Deep Learning for Social Media Fraud and Misinformation Detection.

Table 4: Comparative Analysis of Challenges, Research Gaps, Emerging Trends, and Future Directions in Social Media Fraud and Misinformation Detection

Author	Primary challenges	Research gap	trends	Future directions
Guo and song [46]	Big data scalable computationally efficient and real-time processing	Computational efficient algorithms and context-aware models	GNNs & self-supervised learning	Scalable, flexible and real-time social-media analytics
Alhakami et al. [47]	Algorithmic bias, echo chambers scalability cross-platform disparities	Demand for fairness, flexibility, and quick processing of streaming data	Fairness-aware, Dynamic Memory Networks TCNs multitask learning	Cross-platform, scalable, and principled AI based detection of misinformation

Xia et al. [48]	Lack of transparency and difficult interpretation of misinformation prediction models	Explanatory models combining social-media engagement factors are needed	BERT, LSTM in combination with SHAP-based XAI	Systems to detect misinformation that are trustworthy, interpretable, and feature-aware are what is needed
Sudhakar and Kaliyamurthie [49]	Large-scale misinformation and challenge in validating content on internet	Development of scalable models for vast social-media data analysis	Several ML and deep-learning based methods	Effective large-scale and real-time misinformation identification
Jarrahi and Safari [50]	Publisher credibility and user behavior is not fully exploited	Need to integrate publisher, user, and content-related information	Multimodal publisher-content fusion	Behavior and credibility aware multimodal detection
Singh et al. [51]	AI and human-made misinformation	Generalization of deception across various forms	Ensemble and transformer-based models	Real-time fact-checking and detection of synthetic content
Goyal et al. [52]	Advanced fake profile behavior	Incomplete multimodal account investigation	LSTM-CNN-ANN multimodal learning	Methods to expand behavioral and network-based detection
Tian et al. [53]	Accuracy interpretabilityc omputational cost trade off	No systematic model comparison	Transformer detection	efficient, interpretable model
Sharma et al. [54]	Emotional manipulation and limited dataset	Lack of evaluation of behavior and language use and Restriction of the size of the used dataset in research	Development of a sentiment-aware NLP	Robust dataset and sophisticated behavior modeling.
Sittar et al. [55]	Data imbalance and multilingual data	Overfitting attributable to oversampling	Large language model for data augmentation	Style-aware multilingual synthetic data generation
Xu and Kechadi [56]	False-news detection uncertainty	Benchmark datasets limitations	Fuzzy deep learning	Uncertainty involved detection and better benchmarks

Li et al. [57]	LLM bias, hallucination and misuse	Orther search of reliable detection with LLM	LLM-supported disinformation analyzing	Responsible, ethical and reliable LLMs
Premkumar and Nachimuthu [58]	Privacy and adversarial attacks	Centralized and unimodal constraints	Federated cross-modal graph transformers	Privacy-preserving and adversarially robust systems
Sormeily et al [59]	Constraints on early propagation information	BERTGNN propagation modeling	Highly reliant on large scale propagation data	Early multimodal approach for real-time information detection
Nadeem et al. [60]	Multimedia fabrication and cross-modal inconsistency	Open set and weak visual forgery and semantic analysis	Stylo-semantic fusion	High performance cross-modal forgery detection

4. Research Evolution and Emerging Trends in Multimodal Fraud and Misinformation Detection

Figure 3: Evolution of Multimodal Fraud and Misinformation Detection Approaches on Social Media (2020–2026)

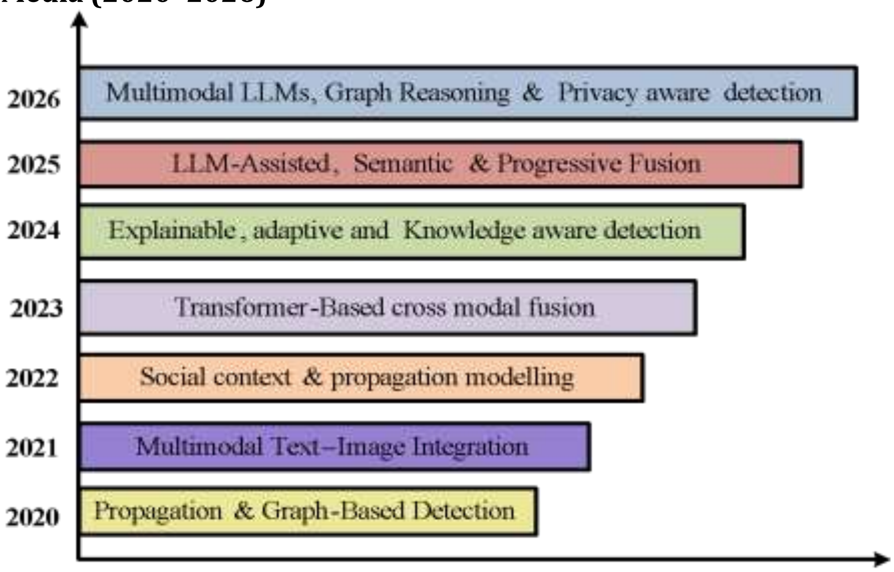
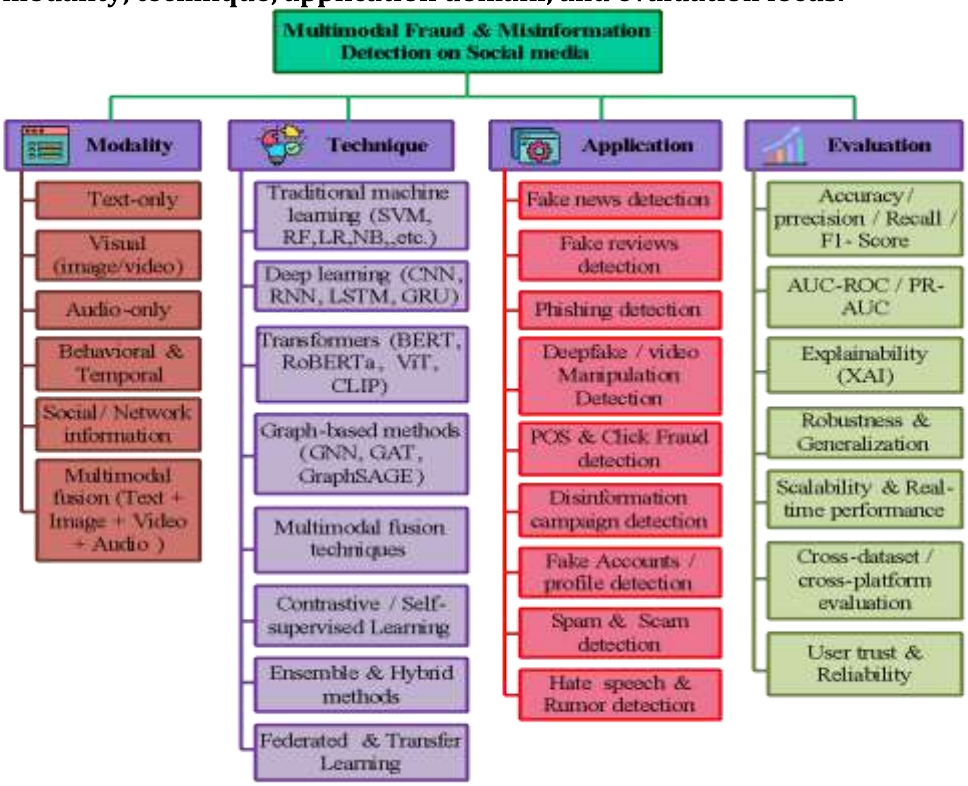


Figure 3 highlights the Evolution of Multimodal Fraud and Misinformation Detection Approaches on Social Media (2020–2026). Recently, research shows that the methods have continued to become more advanced multimodal and context-aware detection features. He et al. [61] emphasized the role of contrastive learning and cross-modal interaction. But, Hamed et al. [62] made use of sentiment and user-emotion data to enhance the misinformation detection process. Goyal et al. [63] reported a trend of deploying ensemble learning for detecting fraudulent user profiles. Meanwhile, Liu et al. [64] made some contributions to multimodal fusion via the hierarchical bilinear method and multimodal consistency, which was mainly used for dealing with joint features of text and image contents. Kishwar and Zafar [65] showed that the detection of localized fake news has changed from traditional machine learning methods to deep learning. These papers together present the emerging areas of cross-modal representation learning, emotional and behavioral analysis, more accurate fusion, ensemble methods, and detection that are sensitive to context.

4.1. Taxonomy of Multimodal Fraud and Misinformation Detection Approaches

Yan et al. [66] employed the combination of CLIP based-contractive and multimodal semantic alignment strategies for the detection of fake news. Bisogni et al. [67] were mainly interested in the detection of deception in audio-visual modality. Singh et al. [68] made use of BERT, visual features, stacked ensemble learning, and performed the task of multimodal fake-news detection. Nwaiwu et al. [69] came up with an interpretable XLM-RoBERTa structure which was capable of handling multi-domain disinformation whereas Katarya et al. [70] presented an overview of several deep-learning methods that were employed for detecting information manipulation. Taken together, these researches indicate that the use of multimodal fusion transformers contrastive learning, ensemble learning methods, and explainable AI can help the identification of fraud and misinformation. Figure 4 shows the Taxonomy of multimodal fraud and misinformation detection approaches based on modality, technique, application domain, and evaluation focus.

Figure 4: Taxonomy of multimodal fraud and misinformation detection approaches based on modality, technique, application domain, and evaluation focus.



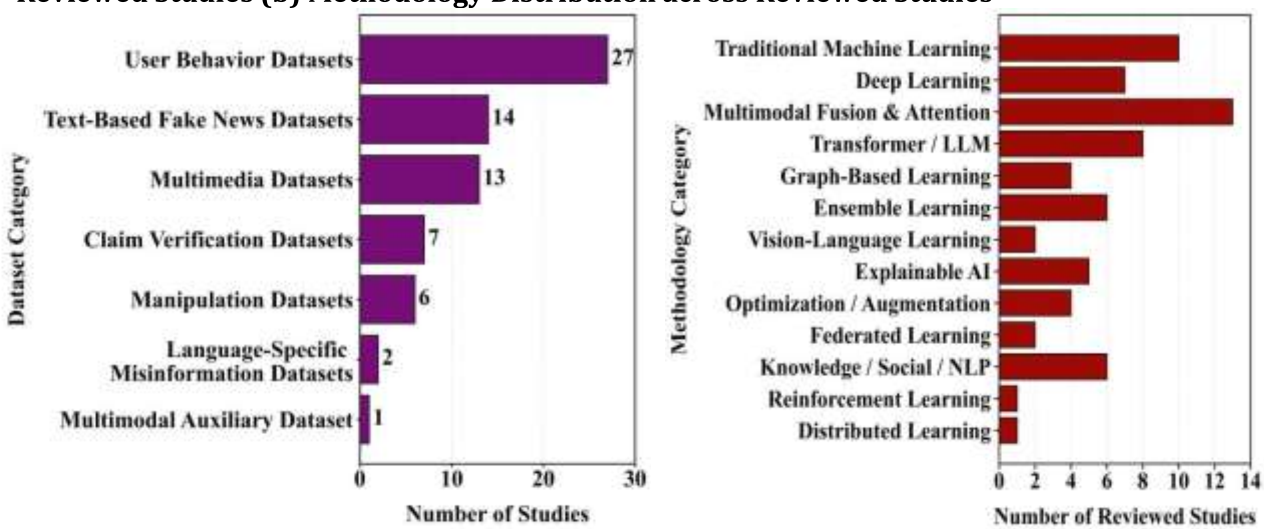
5. Result Analysis

The analysis of reviewed studies reveals a noticeable evolution from conventional machine learning methods through to deep learning, multimodal fusion, transformer/LLM, and vision-language strategies. Twitter, Weibo, and Fakeddit were the top-3 datasets used, and text-image and text-social-context were the predominant multimodal combinations. Research from recent years also showed an increasing number of papers that combined use of attention mechanisms, graph methods, combination of multiple models such as ensemble learning, contrastive learning techniques, explainable artificial intelligent algorithms, optimizations, and federated learning through distributing the training data over multiple devices or computers. In short, findings indicate an emerging trend of combining data of different kinds with each other for better performance in fraud and false information detection.

5.1. Overview of Datasets and multimodal Methodologies in Fraud and Misinformation Detection

Figure 5 show the overview of the fraud and misinformation detection in reviewed studies datasets and detection methodology techniques. Figure 5 (a) displays the dataset classes can be user activity, written fake news multimedia claim checking deception specific language misinformation, and supplemental multimodal datasets, which are a mix of the various forms of information - text image social, and behavior. Figure 5 (b) shows the methodological perspective, the works reviewed here encompass conventional machine learning, neural networks, multimodal fusion and attention, transformers/LLMs, graph-based learning, ensemble methods, visual-linguistic networks XAI optimization approaches, and federated learning. The findings point out that multimodal fusion and attention technique is the method that was mostly adopted. This reflects a gradual change from the old machine learning methods to the new ones which are multimodal and transformer-based Mostly in fraud and misinformation detection scenarios.

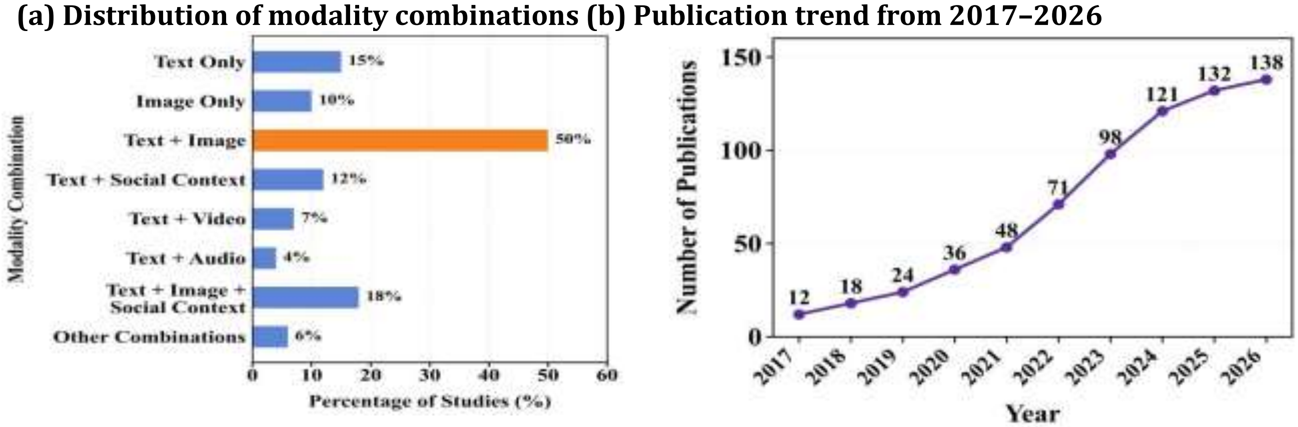
Figure 5: Overview of Datasets and multimodal Methodologies (a) Dataset Distribution across Reviewed Studies (b) Methodology Distribution across Reviewed Studies



(a) **(b)**

5.2. Research landscape and evolution of multimodal fraud and misinformation detection

Figure 6: Research landscape and evolution of multimodal fraud and misinformation detection (a) Distribution of modality combinations (b) Publication trend from 2017–2026



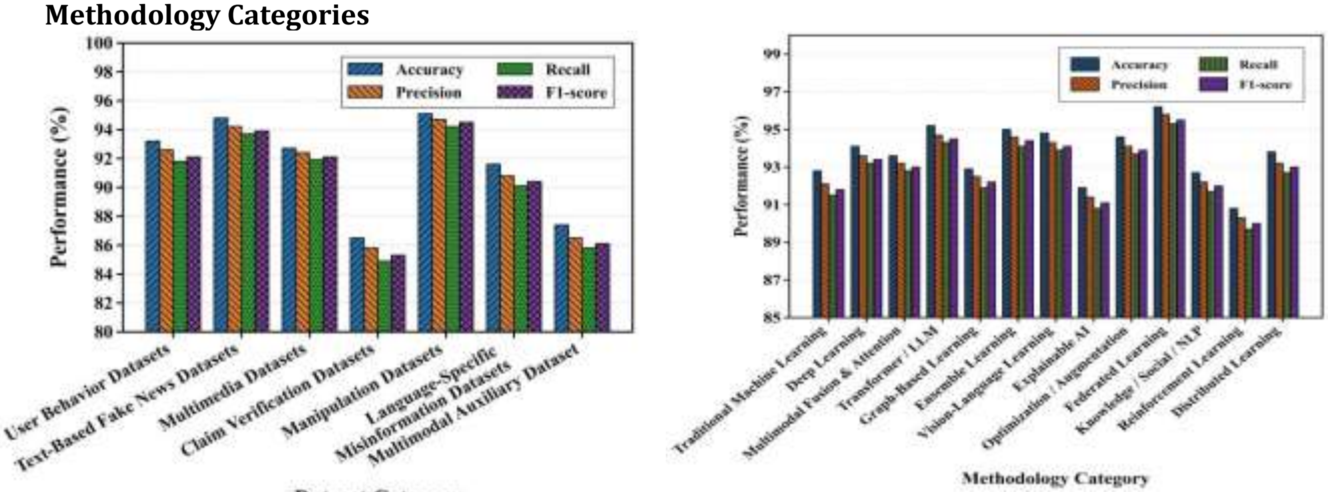
(a) **(b)**

Figure 6 summarizes the research landscape and development of fraud and misinformation detection that use multiple modalities and highlights the most used modality. Figure 6 (a) illustrates that almost all studied papers employed text-image fusion. Also, there is a gradual inclusion of social context information to get semantically and visual. Figure 6 (b) depicts the overall publication trend over this period showing first slow growth and then a marked increase in the research efforts in the past few years. That reflects the research community's turning attention to the advanced deep learning methods, including attention mechanisms and transformers, for combating the problems of fraud and misinformation.

5.3. Performance Comparison of Dataset and Methodology Categories in Multimodal Fraud and Misinformation Detection

Figure 7 displays a comparison of metrics such as Accuracy, Precision, Recall, and F1-score against different dataset and methodology approach categories. In figure 7(a) the overall performance level is relatively higher when using Manipulation Datasets and Text-Based Fake News Datasets. In figure 7 (b) while using methods to produce Claims Verification and Multimodal Auxiliary datasets, the overall performance is relatively lower. In the different ways, observe that Federated Learning, Graph-Based Learning, and Ensemble Learning display relatively higher levels of performance while in contrast optimization, augmentation and explainable AI demonstrate really lower values. Looking at the results in overall performance is likely that advanced multimodal graph-based, ensemble and federated techniques can cause a stronger detection performance of fraud and misinformation.

Figure 7: Comparative analysis of Dataset and Methodology Categories in Multimodal Fraud and Misinformation Detection (a) Performance across Dataset Categories (b) Performance across Methodology Categories



(a) (b)

6. Research gap

A comparative analysis of the current research shows a major research gap in coming up with a single unified way of representing multimedia content for the detection of fake online information and fraud in a community-driven setting. The methods commonly used these days mainly combine various forms of data in small packages, like text-image [20], text-social context [17], or text plus external knowledge and user interaction [42]. Meanwhile, some models have been almost exclusively focused on text [19]. Also, fake profiles [13] and fake news [16] have traditionally been two completely separate concerns, rather than being treated together by capturing the interactions between users and their misleading content. Although some of these techniques are multimodal graphs for learning, or the information that gives context to one another [21]. It has been a universal system that includes user content, user activities, social ties, outside information, and the timeline. It is this fragmentation of ideas that opens up a major research problem for the existing multimodal fraud and misinformation detection literature.

6.1. Challenges

Identification of fraud and misinformation is one of the major challenges when dealing with social media data. This data type is composed of very diverse elements: text imagery the use of hash tags, and even changes in typing tempo - all different factors contributing to the overall picture to be conveyed [1]. To detect fraud or misinformation one has to be able to understand and make sense of all these different types of input effectively [3]. The datasets commonly used are not very rich in content which may hinder a model's ability to detect misinformation patterns from various perspectives [8]. And, user activity is changing and users interact very differently from platform it is hard to know for sure whether an action is just or fraudulent [13]. Incorporating the understanding and interpretation of multilingual content or languages used in one context or another is a further challenge, as one has to represent and classify the meaning behind words accurately [19]. Even more tricky is a combination of sarcasm, emotional persuasion, and confusing contexts which together make it really hard to grasp the actual message conveyed [34]. Also, it is still difficult to combine information and knowledge from external knowledge with the analysis of a social media post, as the evidence for a claim may be coming from very different information sources [42]. Finally, the processing of massive amounts of social-media data with different modalities results in scalability challenges [18].

7. Conclusion

These reviews highlight multimodal deep learning as an effective method to detect frauds and fake news on social media. This examined different aspects of multimodal deep learning, including different datasets, modalities, and methods as well. Based on the review, there has been a shift from traditional machine learning to deep learning approaches, mostly in the areas of multimodal fusion, attention mechanisms, Transformers/LLMs, graph-based learning, ensemble methods, vision-language models, explainable AI, optimization, and federated learning. Of all the studies it came across, those that combined text-image and text-social contexts were the most frequent, with Twitter, Weibo, and Fakeddit being the most widely used sources. The analysis shows that graph-based, ensemble, and federated approaches performed well on average, while multimodal and fusion, attention-based, etc. methods became the leading ones. But modality integration datasets evaluation settings and application scenarios in existing works are still highly fragmented. The research gap pinpointed concerns in particular a unified structure that can simultaneously model content, user-behavior, social networks, external knowledge, and temporal contexts both for fraud and misinformation detection. In general, this review shows that multimodal deep learning offers a sound approach for recognizing the ever-escalating forms of deception in both content and behavior on social media, yet further research is needed for these technologies to develop into trustworthy and efficient methods for social media fraud detection.

Future scope

Future investigations could center on building unified multimodal systems which collectively exploit text image sound behavior social-network, and time-dependent information for more complete identification of fraudulent and misleading content. It is likely that efforts are really increase around detecting cross-platform and fraud with a special focus on under-sourced languages and novel social platforms. The joint employment of advanced transformer/LLM models, graph neural networks, contrastive learning, and multimodal attention should allow better representation of multimodal and deeper contextualization of cross-modal information. Future models could also include instant and early warning techniques to detect false information before a large-scale dissemination as well as models that are not too heavy to deploy at scale. Also, making AI explainable and trustworthy as well as dealing with the problem of privacy, uncertainty, and protection from AI-generated fake material are the main themes through which one might build dependable real-world detection systems. It's through these areas that the transition from single-model, dataset-specific systems can take place and that the field ends up with generalized, scalable, and secures multimodal detection architectures.

REFERENCES

1. Jing, J., Wu, H., Sun, J., Fang, X. and Zhang, H., 2023. Multimodal fake news detection via progressive fusion networks. Information processing & management, 60(1), p.103120.
2. Shu, K., Mosallanezhad, A. and Liu, H., 2022. Cross-domain fake news detection on social media: A context-aware adversarial approach. In Frontiers in fake media generation and detection (pp. 215-232). Singapore: Springer Nature Singapore.
3. Peng, X. and Xintong, B., 2022. An effective strategy for multi-modal fake news detection. Multimedia Tools and Applications, 81(10), pp.13799-13822.
4. Dadgar, S. and Neshat, M., 2022. A novel hybrid multi-modal deep learning for detecting hashtag incongruity on social media. Sensors, 22(24), p.9870.
5. Goyal, B., Gill, N.S., Gulia, P., Prakash, O., Priyadarshini, I., Sharma, R., Obaid, A.J. and Yadav, K., 2023. Detection of fake accounts on social media using multimodal data with deep learning. IEEE Transactions on Computational Social Systems.
6. Mohawesh, R., Obaidat, I., AlQarni, A.A., Aljubailan, A.A., Al-Shannaq, M.A.A., Salameh, H.B., Al-Yousef, A., Saifan, A.A., Alkhushayni, S.M. and Maqsood, S., 2025. Truth be told: a multimodal ensemble approach for enhanced fake news detection in textual and visual media. Journal of Big Data, 12(1), p.197.
7. Dellys, H.N., Mokeddem, H. and Sliman, L., 2025. On the integration of social context for enhanced fake news detection using multimodal fusion attention mechanism. AI, 6(4), p.78.
8. Nistor, A. and Zadobrischi, E., 2022. The influence of fake news on social media: analysis and verification of web content during the COVID-19 pandemic by advanced machine learning methods and natural language processing. Sustainability, 14(17), p.10466.
9. Luo, J., Zhu, P., Wang, Y., Xiao, Z., Li, J., Kong, X. and Zhan, Y., 2026. A Data-Driven Multimodal Method for Early Detection of Coordinated Abnormal Behaviors in Live-Streaming Platforms. Electronics, 15(4), p.769.
10. Mishra, L.N. and Senapati, B., 2026. FraudFusion: A multi-modal ensemble framework for anomaly-based fraud detection in retail POS systems. IEEE Access, 14, pp.10436-10456.
11. Rawat, R., Dibouliya, A., Chaturvedi, N., Rawat, A., Rawat, H. and Dangi, C.S., 2026. A data analytics based federated biometrics for deepfake fraud detection in financial multimedia systems. Discover Computing, 29(1), p.215.
12. Kavya, S. and Sumathi, D., 2025. Multimodal and temporal graph fusion framework for advanced phishing website detection. IEEE Access, 13, pp.74128-74146.
13. Aditya, B.L. and Mohanty, S.N., 2023. Heterogenous social media analysis for efficient deep learning fake-profile identification. IEEE Access, 11, pp.99339-99351.
14. Che, H., Pan, B., Leung, M.F., Cao, Y. and Yan, Z., 2023. Tensor factorization with sparse and graph regularization for fake news detection on social networks. IEEE Transactions on Computational Social Systems, 11(4), pp.4888-4898.
15. Sharma, D.K. and Garg, S., 2023. IFND: a benchmark dataset for fake news detection. Complex & intelligent systems, 9(3), pp.2843-2863.
16. Uppada, S.K., Manasa, K., Vidhathri, B., Harini, R. and Sivaselvan, B., 2022. Novel approaches to fake news and fake account detection in OSNs: user social engagement and visual content centric model. Social Network Analysis and Mining, 12(1), p.52.
17. Park, M. and Chai, S., 2023. Constructing a user-centered fake news detection model by using

classification algorithms in machine learning techniques. IEEe Access, 11, pp.71517-71527.
18. Altheneyan, A. and Alhadlaq, A., 2023. Big data ML-based fake news detection using distributed learning. Ieee Access, 11, pp.29447-29463.
19. Guo, Z., Zhang, Q., Ding, F., Zhu, X. and Yu, K., 2023. A novel fake news detection model for context of mixed languages through multiscale transformer. IEEE Transactions on Computational Social Systems, 11(4), pp.5079-5089.
20. Uppada, S.K., Patel, P. and B, S., 2023. An image and text-based multimodal model for detecting fake news in OSN's. Journal of Intelligent Information Systems, 61(2), pp.367-393.
21. Zhou, Y., Pang, A. and Yu, G., 2024. Clip-GCN: An adaptive detection model for multimodal emergent fake news domains. Complex & Intelligent Systems, 10(4), pp.5153-5170.
22. Gedara, T.M., Gaeta, R., Tomasiello, S. and El Shawi, R., 2025. Toward explainable multimodal fake news detection: A fusion of prompt learning and automated machine learning. IEEE Access, 13, pp.188150-188169.
23. Mbaziira, A.V., Sabir, M.F. and Al Harrack, M., 2022. Lying trolls: Detecting deception and text-based disinformation using machine learning. In Cybersecurity and Cognitive Science (pp. 321-338). Academic Press.
24. Nikumbh, D. and Thakare, A., 2025. Unique multimodal fusion network for fake news detection in social media with attention mechanism. Multimedia Tools and Applications, 84(38), pp.46581-46623.
25. Mouratidis, D., Kanavos, A. and Kermanidis, K., 2025. From misinformation to insight: Machine learning strategies for fake news detection. Information, 16(3), p.189.
26. Seddari, N., Derhab, A., Belaoued, M., Halboob, W., Al-Muhtadi, J. and Bouras, A., 2022. A hybrid linguistic and knowledge-based analysis approach for fake news detection on social media. IEEE Access, 10, pp.62097-62109.
27. Al-Alshaqi, M., Rawat, D.B. and Liu, C., 2025. A BERT-based multimodal framework for enhanced fake news detection using text and image data fusion. Computers, 14(6), p.237.
28. Raza, A., Munir, K. and Almutairi, M., 2022. A novel deep learning approach for deepfake image detection. Applied Sciences, 12(19), p.9820.
29. Zhao, R., Zhu, Y., Heenkenda, S., Wang, J., Abbas, G., Ali, N.A., Elrashidi, A. and Touti, E., 2026. A deep reinforcement learning-based multimodal framework for credible health information mining on social media. Journal of King Saud University Computer and Information Sciences.
30. Abdullah, M., Zan, H., Javed, A., Sohail, M., Mamyrbayev, O., Turysbek, Z., Eshkiki, H. and Caraffini, F., 2026. A multimodal ensemble-based framework for detecting fake news using visual and textual features. Mathematics, 14(2), p.360.
31. Francis, M.A., Kurup, A.R., Premjith, B. and Chakravarthi, B.R., 2025. Multimodal Fake News Classification in Tamil Using Fact-Checked Social Media Content and Cost-Sensitive Learning. IEEE Access.
32. Bhukya, R.K., Shukla, A. and Singh, N., 2026. Multimodal social media fake news detection using RoBERTa and vision transformer encoders with reliability aware adaptive fusion. Discover Computing, 29(1), p.405.
33. Zamil, Y.K. and Charkari, N.M., 2023. Combating fake news on social media: A fusion approach for improved detection and interpretability. IEEE Access, 12, pp.2074-2085.
34. Hajli, N., Saeed, U., Tajvidi, M. and Shirazi, F., 2022. Social bots and the spread of disinformation in social media: the challenges of artificial intelligence. British Journal of Management, 33(3), pp.1238-1253.
35. Abdulridha, F. and Albaker, B.M., 2024. Non-invasive real-time multimodal deception detection using machine learning and parallel computing techniques. Social Network Analysis and Mining, 14(1), p.97.
36. Hamed, S.K., Ab Aziz, M.J. and Yaakub, M.R., 2025. A data augmentation approach based on various GAN models to address class imbalance in fine-grained multimodal fake news datasets. Computing,

107(1), p.52.
37. Abbas, F. and Taeihagh, A., 2025. A multi-level fusion-based framework for multimodal fake news classification using semantic feature extraction. International Journal of Machine Learning and Cybernetics, 16(9), pp.6531-6560.
38. Wu, H., Li, H., Fang, X., Tang, M., Yu, H., Yu, B., Li, M., Jing, Z., Meng, Y., Chen, W. and Liu, Y., 2025. Mdcnn: A multimodal dual-cnn recursive model for fake news detection via audio-and text-based speech emotion recognition. Speech Communication, p.103313.
39. LekshmiAmmal, H.R. and Madasamy, A.K., 2025. A reasoning based explainable multimodal fake news detection for low resource language using large language models and transformers. Journal of Big Data, 12(1), p.46.
40. Joshi, G., Srivastava, A., Yagnik, B., Hasan, M., Saiyed, Z., Gabralla, L.A., Abraham, A., Walambe, R. and Kotecha, K., 2023. Explainable misinformation detection across multiple social media platforms. Ieee Access, 11, pp.23634-23646.
41. Wang, H., Guo, J., Liu, S., Chen, P. and Li, X., 2025. SEPM: Multiscale semantic enhancement-progressive multimodal fusion network for fake news detection. Expert Systems with Applications, 283, p.127741.
42. Fu, L. and Liu, S., 2023. Multimodal fake news detection incorporating external knowledge and user interaction feature. Advances in Multimedia, 2023(1), p.8836476.
43. Aishwarya, C., Venkatesan, M. and Prabhavathi, P., 2026. Debunking Misinformation Using a Deep Learning Approach. Procedia Computer Science, 282, pp.1313-1322.
44. Shukla, P.K., Veerasamy, B.D., Alduaiji, N., Addula, S.R., Pandey, A. and Shukla, P.K., 2025. Fraudulent account detection in social media using hybrid deep transformer model and hyperparameter optimization. Scientific reports, 15(1), p.38447.
45. Kozik, R., Mazurczyk, W., Cabaj, K., Pawlicka, A., Pawlicki, M. and Choraś, M., 2023. Deep learning for combating misinformation in multicategorical text contents. Sensors, 23(24), p.9666.
46. Guo, Y. and Song, W., 2022. A temporal-and-spatial flow based multimodal fake news detection by pooling and attention blocks. IEEE Access, 10, pp.131498-131508.
47. Alhakami, H., Alhakami, W., Baz, A., Faizan, M., Khan, M.W. and Agrawal, A., 2022. Evaluating intelligent methods for detecting COVID-19 fake news on social media platforms. Electronics, 11(15), p.2417.
48. Xia, H., Islam, N., Zhu, D., Zhang, J.Z., Behl, A. and Roohanifar, M., 2026. A novel method of fusing Bert-LSTM and XAI for social media disinformation identification. Annals of Operations Research, pp.1-25.
49. Sudhakar, M. and Kaliyamurthie, K.P., 2024. Detection of fake news from social media using support vector machine learning algorithms. Measurement: Sensors, 32, p.101028.
50. Jarrahi, A. and Safari, L., 2023. Evaluating the effectiveness of publishers' features in fake news detection on social media. Multimedia Tools and Applications, 82(2), pp.2913-2939.
51. Singh, I., Choudhary, C. and Gupta, P., 2025. Ensemble-based framework for fake news detection in social-media. Journal of Computer Information Systems, pp.1-16.
52. Goyal, B., Gill, N.S., Gulia, P., Alduaiji, N., Shukla, P.K., Shukla, P.K. and Vishnoi, S., 2026. Deep learning with optimization for social media fake profile detection. Social Network Analysis and Mining.
53. Tian, Y., Xu, S., Cao, Y., Wang, Z. and Wei, Z., 2025. An empirical comparison of machine learning and deep learning models for automated fake news detection. Mathematics, 13(13), p.2086.
54. Sharma, D.K., Singh, B., Agarwal, S., Kim, H. and Sharma, R., 2022. Sarcasm detection over social media platforms using hybrid auto-encoder-based model. Electronics, 11(18), p.2844.
55. Sittar, A., Smiljanic, M., Guček, A. and Grobelnik, M., 2026. Fake News Detection Through LLM-Driven Text Augmentation Across Media and Languages. Machine Learning and Knowledge Extraction, 8(4), p.103.
56. Xu, C. and Kechadi, M.T., 2024. An enhanced fake news detection system with fuzzy deep learning.

IEEE Access, 12, pp.88006-88021.

57. Li, T., Yu, J. and Zhang, H., 2022. Web of things based social media fake news classification with feature extraction using pre-trained convoluted recurrent network with deep fuzzy learning. Theoretical Computer Science, 931, pp.65-77.
58. Premkumar, D. and Nachimuthu, S.K., 2026. Privacy-preserving cyberthreat detection in decentralized social media with federated cross-modal graph transformers. Scientific Reports, 16(1), p.3608.
59. Sormeily, A., Dadkhah, S., Zhang, X. and Ghorbani, A.A., 2024. MEFaND: A multimodel framework for early fake news detection. IEEE Transactions on Computational Social Systems, 11(4), pp.5337-5353.
60. Nadeem, M.I., Ahmed, K., Zheng, Z., Li, D., Assam, M., Ghadi, Y.Y., Alghamedy, F.H. and Eldin, E.T., 2023. SSM: Stylometric and semantic similarity oriented multimodal fake news detection. Journal of King Saud University Computer and Information Sciences, 35(5), p.101559.
61. He, Z., Wang, H. and Li, L., 2025. Fake News Detection Based on Contrastive Learning and Cross-Modal Interaction. Symmetry, 17(8), p.1260.
62. Hamed, S.K., Ab Aziz, M.J. and Yaakub, M.R., 2023. Fake news detection model on social media by leveraging sentiment analysis of news content and emotion analysis of users' comments. Sensors, 23(4), p.1748.
63. Goyal, B., Gill, N.S., Gulia, P., Alduaiji, N. and Shukla, P.K., 2025. Instagram fake profile detection using an ensemble learning method. Scientific Reports, 15(1), p.26464.
64. Liu, Y., Bing, W., Ren, S. and Ma, H., 2024. BC-FND: an approach based on hierarchical bilinear fusion and multimodal consistency for fake news detection. IEEE Access, 12, pp.62738-62749.
65. Kishwar, A. and Zafar, A., 2023. Fake news detection on Pakistani news using machine learning and deep learning. Expert Systems with Applications, 211, p.118558.
66. Yan, F., Zhang, M., Wei, B., Ren, K. and Jiang, W., 2024. Sard: Fake news detection based on clip contrastive learning and multimodal semantic alignment. Journal of King Saud University Computer and Information Sciences, 36(8), p.102160.
67. Bisogni, C., De Marsico, M., Loia, V., Miglino, A. and Nappi, M., 2026. Decoding deception: interpretable multimodal models for audio–visual Lie detection. Complex & Intelligent Systems.
68. Singh, P., Srivastava, R., Rana, K.P.S. and Kumar, V., 2023. SEMI-FND: Stacked ensemble based multimodal inferencing framework for faster fake news detection. Expert systems with applications, 215, p.119302.
69. Nwaiwu, S., Jongsawat, N. and Tungkasthan, A., 2025. Decoding disinformation: A feature-driven explainable ai approach to multi-domain fake news detection. Applied Sciences, 15(17), p.9498.
70. Katarya, R., Dahiya, D. and Checker, S., 2022. Fake news detection system using featured-based optimized msvm classification. IEEE Access, 10, pp.113184-113199.