Journal of Al-Farabi for Medical Sciences volume (3), No (3) September (2025)

ISSN: 2957-7675 (Print)
Journal of Al-Farabi for Medical Sciences
https://iasj.rdd.edu.iq/journals/journal/view/96
Issued by Al-Farabi University

A Systematic Literature Review through Comparative and Critical Evaluation

[٢]Zainab khaled Abdullah[1], Ahmed Jassem Mohammed

[1]Computer Science, College of Computer Science and Information Technology, University of Anbar, Email: zai24c1006@uoanbar.edu.iq Ramadi, Iraq,

[٢] Artificial Intelligence, College of Computer Science and Information Technology, University of Anbar, Ramadi, Iraq, Email: a.j.aljaaf@uoanbar.edu.iq

Abstract

The rapid and increasing spread of fake news and rumors in social media platforms has posed issues that undermine the trustworthiness of digital information, which has led to rapid developments of automated verification systems. This work provides a comprehensive and critical evaluation of the literature on false news detecting techniques and systems. The study emphasizes the methodological evolution of the applied approaches, from the conventional statistical models to the latest advanced Natural Language Processing (NLP) techniques with hybrid feature extraction such as textual features, sentiment analysis, and structural propagation dynamics. This review identifies the existing gaps in the research and the limitations of the literature. Most importantly, it addresses the problems of comprehending colloquial dialects that use linguistic metaphors and sarcasm, the problems of generalizing across platforms, and the "black box" nature of sophisticated models. A roadmap focusing on high-precision verification systems is discussed, along with future research prospects. The importance of using Explainable AI (XAI) tools to increase the transparency and trustworthiness of models in real-world settings is highlighted.

Keywords: Fake news, rumor verification, natural language processing, explainable artificial intelligence, systematic reviewer.

.١ Introduction

With recent technological progress and the growing popularity of social media apps, conventional ways of sharing news and information are gradually fading. These traditional methods have largely been replaced by modern communication platforms. This shift is driven by the fast spread of content, easy access, and dynamic online environments. Such platforms actively involve users by providing a space for conversation and opinion-forming, which ultimately changes how people interact with information. According to the Social Media and News Fact Sheet (July 15-August 4, 2024), 54% of U.S. adults indicated that they get their news via social media at least occasionally ,

while smaller shares routinely do so on platforms like Instagram.[١]

Building on this shift, the rise of digital communication has altered not just how information is produced and shared, but also the dynamics of trust and authority online. While these platforms offer new ways to interact and share content openly, they also blur the line between professional journalism and everyday user posts. Without clear editorial boundaries, it becomes much easier for unverified claims, rumors, and fabricated news to spread rapidly, raising serious questions about the reliability of information in the public sphere.

These consequences extend far beyond the digital realm, directly threatening political stability, public trust, and social cohesion. The spread of fake news has demonstrably impacted real-world events, from influencing the 2016 U.S. presidential election [2] to triggering severe public health crises. For instance, during the COVID-19 pandemic in Iran, rumors promoting methanol as a cure resulted in hundreds of deaths and thousands of poisonings [3]Similarly, in Brazil between 2016 and 2018, exposure to vaccine misinformation reported by nearly 90% of surveyed individuals was linked to a noticeable drop in immunization rates [4]. The sheer scale of this crisis is illustrated by the fact that in 2020 alone, media platforms had to remove approximately 50 million misleading posts related to COVID-19.[٥]

Given the massive volume and rapid spread of online disinformation, manual verification is no longer a viable strategy. Consequently, there is a critical need for robust, automated systems capable of analyzing content in real time. Driven by recent advancements in Natural Language Processing (NLP) and machine learning, these computational approaches have become essential for detecting false claims, curbing the spread of rumors, and protecting the integrity of public discourse.[٦]

Recent advancements in the field highlight a shift towards integrating deep linguistic context, sentiment features, and multimodal data. For instance, the JLFND framework successfully outperformed traditional BERT baselines by fusing entity, relation, and stance features. It achieved impressive results, recording an accuracy and F1-score of 0.94 and 0.95 on the PolitiFact dataset, respectively, and reaching 0.98 for both metrics on GossipCop [7]. Furthermore, hybrid architectures that incorporate visual elements have demonstrated significant performance gains. The SSA-MFND model, for example, reached an accuracy of 0.912 and an F1-score of 0.914 on the Weibo dataset, alongside a 0.975 accuracy on English datasets, further confirming the robustness of multimodal verification approaches.[٨]

Several research have analyzed the social, political, and psychological effects of falsified information and have explored various automated detection systems, but there is yet no final solution that can totally prevent or reliably identify all forms of fabricated content. Traditional methods have fallen short, especially with the advent of sophisticated media manipulation tools and AI-created content that closely mimics actual information. the continuous development of false news necessitates additional study and more powerful detection methods for improving accuracy and flexibility in real world scenarios.[٩]

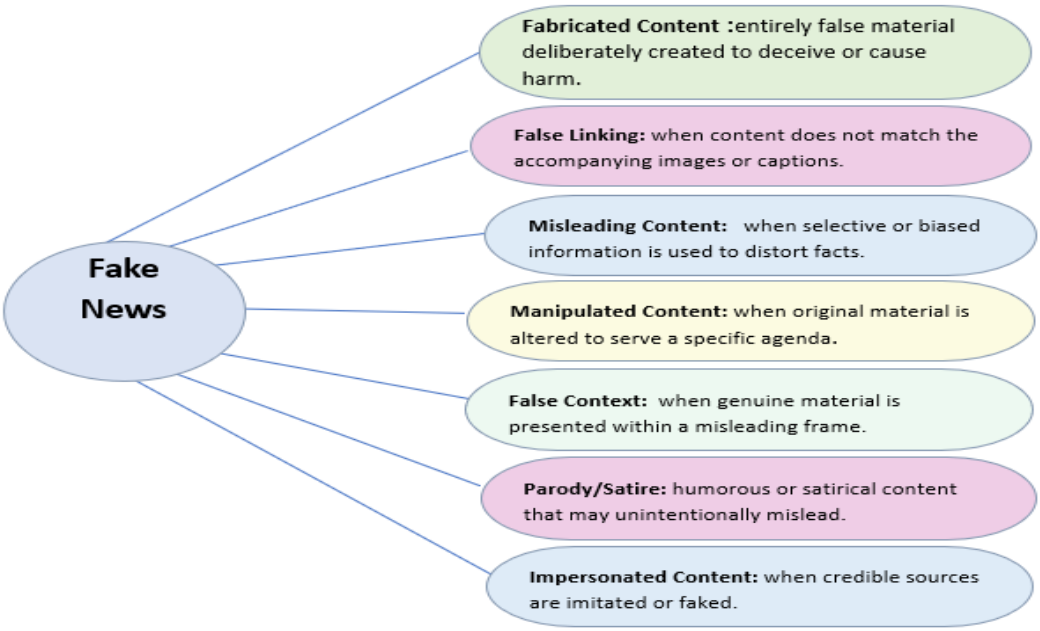
The rest of the paper is organized as follows. In the first part, the issue is introduced with a basic overview. Then the definition of fake news and its types are given. The final section discusses the most critical techniques for false news and content detection. Section four investigates the performance of algorithms employed in research investigations. Section four presents a table summarizing the details of the experiments. The final section discusses the most important obstacles and future research trends. The sixth section discusses ethical difficulties and prejudice for the attention of the researchers. The seventh section is the last to finalize the research.

.ا Type Of Fake News

To systematically analyze the structural variations of deceptive content, contemporary research operates within a framework of 'information disorder'. Within this taxonomy Sandu et al. [10] delineates three overlapping yet distinct concepts:

- Misinformation refers to inaccurate claims shared without malicious intent, often arising from a lack of fact-checking or reliance on unverified sources.
- Disinformation, conversely, is characterized by the orchestrated dissemination of false data. It is typically driven by organized campaigns aiming to manipulate public perception and erode institutional trust.
- Fake news represents a specialized subset within this framework. It involves deliberately fabricated media designed to mimic legitimate news, seeking to deceive audiences, drive engagement, or advance ideological agendas. Such content frequently relies on biased contextual manipulation to incite social polarization.

To systematically classify these structural nuances, Wardle [11] proposes (Figure 1) a typology comprising seven distinct forms of fabricated content:



1. **Fake News Approaches And Techniques**
   Fake news differs from other deceptive content in various ways. It exploits sensationalist headlines, spreads quickly online, and lacks reliable sources [11], [12]. It also uses emotionally charged and polarizing language to incite wrath and fear, strengthening social and cultural divisions [11], [13].

Psychological and social variables influence false news susceptibility as well as language manipulation. Online misinformation is more likely to be believed by people with poor self-esteem and high social influence. Fake news uses more third-person pronouns, shorter sentence structures, less lexical diversity, and more negative emotions [6]. These properties of fraudulent content have been used by researchers to build methods for detecting it .

Deep learning, linguistic and statistical analysis, and context, meaning, and emotion are used to detect fake news as (figure 2). This section describes how academics established accurate models to spot fraud in texts, photos, and social media, from conventional methods to modern intelligent systems.

3.1 Linguistic And Statistical Approaches

In the research of fake news identification ,

various of pure linguistic and statistical aspects are usually used, such as word frequency, TF-IDF scores, n-grams and stylistic indicators. These elements are aimed to capture shallow textual patterns (e.g., the overuse of trigger words, punctuation, or repeating phrase structures) that typically distinguish false texts from real ones. For example, in the work Detecting Fake News from Social Media, the authors used only the textual content from Twitter, based on data from verified and unverified accounts; this purely statistical representation of words was later used to train and evaluate several classifiers [14]. Similarly, another study on misinformation during the COVID-19 pandemic analyzed a dataset of tweets categorized as authentic or fraudulent, utilizing Bag-of-Words (BoW) and TF-IDF representations to direct the analysis [15]. These approaches reveal important limitations while demonstrating the utility of surface-level textual elements. Such models usually lack in modeling the deeper semantic and contextual layers of text which is important for retrieving sophisticated or subtle misinformation. They mostly rely on statistical patterns. Previous research, therefore, have shown that linguistic and statistical models alone provide limited accuracy. Their true worth is realized when combined with semantic representations, when they are much

more effective as supplemental signals rather than as primary detectors [7]. Studies on this area indicate that combining lexical features (such TF-IDF, n-grams, and count-based vectors) with semantic embeddings (like GloVe) results in a much deeper text representation. “The hybrid architectures allow the detection systems to learn the language’s structural patterns and the qualitative aspects, such as emotional tone and informality, that characterize deceptive communication [16].

3.1 Integrating Emotional & Contextual And Semantic Dimension

the analysis of sentiments in news headlines and user comments provides a complementary approach to detecting fake news and understanding media manipulation. While headlines often reflect publishers’ intentions and strategies frequently relying on exaggeration or emotional bias to capture attention user comments represent the audience’s immediate response, indicating levels of acceptance, rejection, or awareness of manipulation. The divergence or alignment between these sentiments thus becomes a key indicator of biased or misleading content. Researchers argue that fake news detection should explicitly integrate emotional and affective cues from both content and audience reactions, as this dual perspective reveals manipulation strategies more effectively than content analysis alone. Fake news commonly relies on negative emotions such as fear, disgust, and anger to attract attention, whereas real news tends to be more balanced or positive, eliciting trust and anticipation. In the study, headline sentiment was analyzed using TextBlob, while audience comment sentiment was classified with NRCLEx, demonstrating the potential of this combined method to enhance the reliability of fake news detection systems [17]. this dual-layer approach underlines the connection between the purpose of the communication and the perception of the audience. It provides an all-inclusive knowledge of the dynamics of disinformation. Studies that use tools such as TextBlob (sentiment analysis of the headlines) and NRCLex or VADER (reactions of the audience) have shown that combining emotional signals from the publisher and reader improves the performance of fake news classifiers. Moreover, non-verbal affective cues like emojis are also utilized in the sentiment modeling process since these symbols often express subtle emotions like sarcasm, laughing or fury that cannot be captured through textual analysis alone [12].

This approach highlights the prevalence of negative affect in disinformation, but most studies emphasize the role of negative emotions (e.g., fear, wrath, and disgust) in the spread of bogus news to increase its virality. One study, however, disputed this agreement by demonstrating that bogus news can also use pleasant emotions . This view stresses that misleading information can be intentionally created to generate optimism, hope, or even laughter to increase believability, decrease resistance, and ease dissemination among like-minded people. This difference further highlights the benefit of capturing the complete range of emotion in detection models as opposed to only negative sentiment. To handle this complexity, one study concatenated the textual features obtained through TF-IDF and the emotional features obtained through sentiment analysis to capture manipulation strategies, and also incorporated the domain-specific lexical features, such as Thai herbal and disease names, to identify health-related fake news [18].

According to study, fake news detection algorithms work better when emotional, contextual and semantic levels are considered combined. The combination allows models to mimic how phony emotions might support manipulation through social structures and verbal messages. Word sentiment, network centrality, semantic similarity and message volume were predictive [19]. For example, one paper suggests a full feature engineering approach with 39 sentiment, linguistic and named entity variables. This multi-dimensional architecture will provide an opportunity for deep learning models like GRU, LSTM and RNN to understand lexical regularities, emotional polarity and entity specific patterns to recognize manipulation methods better. The emotional and linguistic cues complement each other to provide subtle deception cues and the diversity of features is required for false news identification [2]. In later works, there has been a gradual movement towards the combination of classical textual elements with deeper semantic and contextual models. This progression is exemplified by approaches like Named Entity Recognition (NER), Relationship Classification (RFC) and Stance Detection (SD) that move the analysis focus from superficial lexical frequencies to the pragmatic and

semantic structures underneath news content. NER finds the major entities and events, RFC finds the relational relationships among them and SD finds attitudes towards concrete claims or players. The integration of these methodologies provides a unified analytical scheme to enhance the interpretability and depth of false news identification. A prominent progress in this line of research is the Joint Learning Framework for Fake News Detection (JLFND) . It uses an improved BERT-based architecture to jointly learn NER, RFC and SD in a multi-task learning framework. The model adopts hierarchical attention methods to capture local and global dependencies, so as to improve the ability to evaluate factual coherence and semantic consistency [7].

Although integrative approaches help to get a fuller picture of disinformation, emotional expressions are different for different languages and platforms and contextual cues may be different as social interactions are different. The models need to be improved, especially with adaptive architectures that can grasp shifting semantic and emotional conditions in real time.

3.2 Temporal And Network-Based Diffusion Dynamics

Researchers have begun to study the dynamics of spreading bogus news. Empirical studies have shown that fake news usually has a burst-like diffusion pattern: it spreads with an accelerated speed in the initial phase, due to novelty and emotional resonance, quickly getting shares and interactions, but then falls off sharply in engagement after a relatively short period of time. This temporal pattern sets it apart from genuine news, which tends to disperse more evenly and keep interest over longer durations. To capture these dynamics, researchers use diffusion tree analysis, temporal modeling, and network centrality measurements, among other methods. on both cases, the methods find that, on general, falsehoods spread to more people, reach a sharper peak and disintegrate more swiftly than truth-based stories. This unique diffusion curve, a marker of untruth, of quick acceleration followed by fast decrease, has been consistently observed in large scale social media datasets [20]. Earlier studies have criticized the generalizability of cascade analysis and network measurements across platforms and contexts . Diffusion based models are less predictive in real time as they observe patterns of spread afterwards. Datasets with short temporal coverage are often affected by outdated information or broken links, which makes replication difficult and undermines current approaches. To model the complexity of disinformation dissemination, scholars propose deep graph modeling with temporal and sentiment supervision [21]. Extending this line of work it has been proposed to use spectral graph deep learning to detect the structural properties of the relationships between news items, and to combine it with Gated Recurrent Units (GRU) to represent the temporal dynamics of information flow. Their results show that the combination of the network structure and temporal sequence can considerably increase the accuracy of prediction and enable more efficient identification of fake news than existing methods [22]. In addition to analytical models based on networks, some research have also addressed behavioral measures to reduce the spread of fake news. For example, researchers examined the influence of crowd ratings versus expert ratings on users' impression of news reliability and inclinations to spread the news . The results demonstrated that the inclusion of ratings, either by experts or ordinary users, diminishes the perceived credibility of false material and inhibits its re-dissemination. This shows the need to integrate social and behavioral aspects with technical models to increase the efficiency of fake news identification [23].

3.3 Integrating Textual And Visual Cues

multimodal models that integrate text and visuals have emerged as a potential path to improve detection accuracy . A lot of bogus content uses random/unrelated photographs or pictures out of context. From a semiotic perspective, the congruence of word and image creates credibility, while incongruence indicates manipulation. Cognitively, when words and visuals are not congruent, it is harder for readers to retain and comprehend information, and they lose trust.

In technical terms, models like CLIP learn to construct joint spaces of representations for text and images so that the system can estimate how well they match. Based on this, recent work has suggested multimodal models with Co-Attention processes that assess semantic

consistency at two levels low-level entities such as people, locations, and organizations and high-level descriptive alignment between text and images. These approaches have shown a high potential in spotting inconsistencies as deceptive news often breaks this consistency [8]. Apart from these technical contributions, the authors also propose a new multimodal dataset for Bengali comprising text and image data over four categories, viz., factual news, misinformation, rumors, and misleading headlines, providing an improved empirical basis for multimodal false news detection [24]. Besides simple semantic consistency, scholars have also noted the relevance of picture manipulation analysis. Many altered or compressed photos have unique patterns which are visible in the spatial and frequency domains. For example, the spatial features can be used to describe the visible content such as objects and events, while the frequency features based on the DCT-CNN can expose the compression artifacts, watermarks or splicing as the frequent falsification signs. Models can combine these visual signals with textual information to better identify discrepancies and enhance classification performance [25]. More recently, the findings have been extended by employing enhanced CLIP models and vision architectures such as Vision Transformers (ViT) to extract broad visual properties. The techniques can be used to assess the coherence between text and images, but also to allow the systems to detect manipulations, such as the reuse of old photos or changes in the visual environment. Computer vision has the potential to use visual and textual information to significantly improve the reliability and accuracy of detecting fake news, which is beyond the reach of language analysis [26]. Finally, some academics have suggested fully integrated multimodal frameworks in which many approaches are combined within a single decision-making system. One study employed a methodology combining NLP and LSTM for text analysis, computer vision and ANFIS for detecting image manipulations, and fuzzy matching for text-image consistency checks. These outputs were combined in a fuzzy inference system, which gave a better accuracy and interpretability in the detection process [27].

Recent work has concentrated on more extensive hybrid frameworks based on multimodal advancements combining mostly textual and visual information. These approaches incorporate several kinds of features including linguistic, semantic, emotional, network, and transfer-based representations to build integrated detection models that can capture misinformation from multiple analytical angles.

3.4 Unified Multi-Feature Frameworks

Another work further extends the integration of different feature types and applies a combination of textual, semantic, and sentiment-based features for detecting fake news content. The features used Bag of Words, TF-IDF, and Hashing which are statistical representations of the tweets. Semantic features were included by word embeddings and LSTM representations, so that the model can capture contextual meaning. The authors also included sentiment based features using the Sentiment Majority Voting Classifier (SMVC) that labeled each tweet as good or negative. Finally, the study presented hybrid transfer features by mixing the outputs of Decision Tree and LSTM models which leads to richer feature representations [28].

the examined studies indicate the fact that no single strategy is sufficient for reliable false news detection. Diffusion-based, textual, semantic and sentiment-centered techniques provide useful insights, but their limits become apparent when employed in isolation. The most promising trend is in hybrid and multimodal models that include linguistic, cognitive, structural and affective characteristics hence providing more robust and generalizable answers.

(Table 1) summarizes the different approaches used in automated systems, highlighting the benefits and limitations of each.

Table 1. Comparative Overview of Fake News Detection Approaches

Approach	Key Characteristics & Strengths	Limitations
Linguistic and Statistical	Captures shallow textual patterns like word frequency, TF-IDF, and stylistic indicators.	Lacks deep semantic and contextual modeling; limited accuracy when used alone.
Emotional, Contextual & Semantic	Integrates sentiment cues from publishers and audiences; effectively reveals manipulation strategies.	Emotional expressions and contextual cues vary significantly across different languages and platforms.
Temporal and Network-Based	Captures the accelerated spread and sharp peak characteristic of fake news diffusion.	Less predictive in real-time; relies on post-spread cascade analysis and network measurements.
Multimodal (Textual + Visual)	Assesses semantic consistency between words and images; exposes image manipulations or compression artifacts.	Higher architectural complexity and relies on computationally heavy tools like CLIP or ViT.

From the above table, we can see that no one method can perfectly detect fake news in all cases. Linguistic and statistical approaches are semantically shallow and sentiment-based approaches are not applicable to cross-lingual and platform variation. Furthermore, network based approaches are not capable of early detection and multi-modal models are computationally expensive.

This means that one approach is no longer sufficient on its own. This emphasizes the significance of hybrid frameworks that combine several techniques to overcome the limitations of each one of them and to achieve higher accuracy and robustness.

4. Evolution And Comparative Analysis Of Fake News Detection Models

The detection of fake news has transitioned from statistical and classical machine learning techniques to deep learning and transformer-based models. Previous works showed reasonable performances, but they suffered from semantic understanding and domain adaption. Recent advances with enriched representations and hybrid architectures have dramatically improved robustness and precision. This section contains a statistical study of the classification approaches described in the literature, with a focus on major methodological patterns (Figure 3).

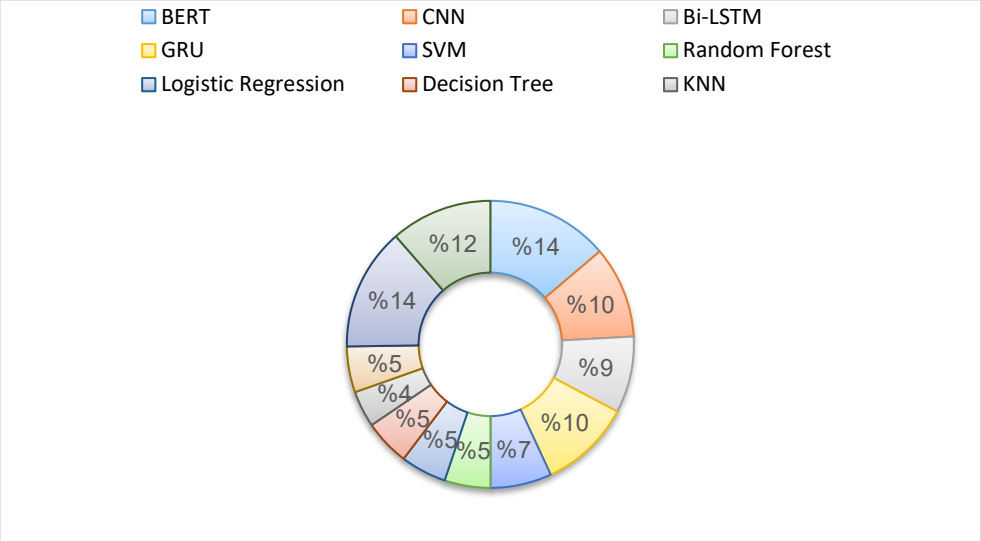


Figure 1. A statistical study of the classification algorithms

4.1 Statistical With Classical Ml

Having described the many techniques that researchers use to detect false news, it is important to study how these models work in the real world. Comparative evaluations show that typical machine learning models based on surface features have middling success, whereas deep learning architectures enhanced with semantic, sentiment, or multimodal variables routinely outperform them. Understanding the reasons for these performance disparities highlights the strengths and limitations of each modeling paradigm, and suggests future study directions.

One of the central concerns in fake news detection is why do relatively simple models such as SVM occasionally beat more complex deep learning architectures such as LSTM? The main one is the nature of the textual representation. When raw text is translated into accurate statistical characteristics, e.g. using the TF-IDF, then the surface-level lexical indications such as word frequency and distribution can be adequate for effective categorization. SVM is well known for its capacity to handle high dimensional and sparse feature spaces as well as to find a stable and optimal decision boundary.

This seeming gain should not be mistaken for a fundamental advantage of shallow models over deep ones. It is generally due to limited dataset size and breadth. Simple lexical or sentiment-polarity signals allow shallow models to easily detect bogus news from authentic news. Deep learning algorithms perform best on large and diverse datasets and they employ richer semantic and contextual frameworks. Smaller or domain ـ

specific corpora are not making use of their representational capability and are either underperforming or overfitting.

This argument is corroborated by a study on Twitter sentences where SVM outperformed other classifiers clearly with 98% accuracy, whereas LSTM attained accuracy values ranging
from 54% to 65% [14]. These results emphasize a paradox: the apparent robustness of classic machine learning models is not due to a more profound representational power, but rather to their compatibility with restrictive data contexts that make the classification task deceptively simple. A similar type of evidence was given by a study that addressed a fake-news detection on social media by a text-based approach integrating word-embedding techniques and both machine learning and deep learning models . The authors use Truth-Seeker dataset and extract features using TF-IDF, Word2Vec and Fast-Text after performing a pre-processing pipeline of text cleaning and tokenization. Results indicated that SVM with TF-IDF achieved the greatest accuracy 99% while CNNs performed well with semantic embeddings but did not top SVM. This study further supports the idea that detection performance is mostly determined by the match between feature representation and model architecture rather than the model choice alone [6]. Another work made a similar finding where hybrid transfer features were produced by merging outputs from Decision Tree and LSTM models with typical textual representations like Bag-of-Words, TF-IDF,

Hashing and embeddings. Surprisingly, Logistic Regression on these transfer features had the highest accuracy (98.9%) which surpassed deep learning approaches such as LSTM (93%) [28]. This highlights the notion that given sparse data, well-engineered feature based classical models can outperform deep learning models that need big and diverse corpora to unleash their full potential.

Data patterns were also crucial in assessing the performance of algorithms. The fake news was characterized by recurrent and explicit linguistic traits, such as sensational words and short constructions that drew samples closer together in the feature space. In contrast, the larger stylistic and semantic variation of genuine news led to a higher spread among the samples. This means that the effectiveness of algorithms depends on how homogeneous the data is, and can explain the advantage of particular machine learning models in certain environments. One study utilized several algorithms (Logistic Regression, K-Nearest Neighbors (KNN), and Multilayer Perceptron (MLP) to a Kaggle dataset of around 21,000 pre-labeled news items fraudulent vs real. All models obtained high accuracy (>95%), with KNN being excellent in fake news detection, while Logistic Regression and MLP provided a balanced performance across both classes [29]. But the reliance on repetitious domain data points to the possibility of overestimating resilience.

Limitations of conventionalfake-news categorization methods hinder their utility. First, the textual noise severely hinders their ability to extract semantic features. Second, they do not alleviate the class imbalance that degrades minority accuracy. Third, assumptions on simple textual data are applied. Such challenges highlight the need for more sophisticated models that can deal with noise, imbalance and language complexity. For the numerically represented WELFake dataset utilizing TF-IDF three ensemble models were used: 1) Balanced Random Forest for imbalance, 2) XGBoost as an efficient gradient-boosting method and 3) Light-GBM as fast training and efficiency. The ensemble techniques consistently outperform standard models in earlier investigations. The top performing models were Light-GBM (97.6% accuracy and 97.7% F1-score), XG-Boost (96.8%) and Balanced Random Forest 94.4% [30.[

The majority of the prior work was limited to a single domain, such as politics, leading to models that are not very successful for other news topics due to their narrow scope. To overcome this constraint, the present study used multi-domain data comprising politics, society, economy, health, and entertainment news to improve generalization. After preprocessing and numeric representation, a public textual dataset from Kaggle was used to apply multiple classifiers, such as Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), K-Nearest Neighbors (KNN), and Support Vector Machine (SVM). The performance of models varied and the most accurate was SVM with 93.61% accuracy, 91.32% precision and 95.70% recall [31]. Statistical methods are limited to shallow linguistic patterns and cannot verify factual consistency or capture deeper semantic linkages. This has led to the inclusion of external knowledge signals that enable models to check assertions, find contradictions in entities, and assess the plausibility of material. When paired with verbal or affective aspects, knowledge-based signals are substantially more accurate and generalizable. For example, the authors of [20] used eight linguistic and knowledge-based attributes, such as source credibility, topic coverage, and fact-checking indicators, and found that the accuracy achieved by the linguistic features alone was 77.6% and 89.46%, knowledge-based features alone was up to 81.2%, and combining both features was 94.4% by using Random Forest.

These findings show the efficiency of hybrid knowledge-enhanced methods on fake-news identification. Empirical studies give significant evidence that the use of features based on sentiment and lexicons in the modeling phase can close this gap. For example, the Emo-SL Framework presented for Arabic sentiment analysis revealed that the performance of the models is significantly improved when affective cues from emojis are combined with textual data. The study used around 58,000 Arabic tweets and found that the accuracy of SVM classification rose from 70% for text-only input to 89% when integrating emoji sentiment ratings with lexical information, with an F1-score of 0.89. This result is in line with the broader evidence that affective signals, either conveyed by textual polarity, domain-specific lexicons or visual-emotive cues, are significant to improve classification robustness across several NLP tasks, including fake-news detection [32]. A study on a Thai-language health dataset, based on this evidence, showed that the addition of sentiment analysis and health-specific lexical cues in TF-IDF representations improved model accuracy by more than 7% compared to text-only approaches, with XGBoost providing the strongest baseline performance.[١٨]

(Table 2) summarizes classical and statistical machine learning models.

Table 1. classical and statistical machine learning models

REF/ Year	Language	Features represent-tation	Detection Method	Data Set	Performance	Limitations
[18] 2024	Thai	semantic Representat-ions	Machine Learning	health dataset collected from online platforms	Accuracy of XGBO = 0.887, F1-score = 0.885	-Limited domain. -imbalance between fake/real news.
[6] 2024	English	Statistical, semantic Representat-ions	Deep Learning + Machine Learning	TruthSeeker Dataset	SVM + TF-IDF = 99% Accuracy, F1 = 99%; CNN-3 + TF-IDF = 98.99%	-Single-domain dataset. - Ethical aspects not addressed.
[14] 2024	English	Statistical, Representat-ions	Deep Learning + Machine Learning	Twitter /X dataset	Accuracy of SVM: 98% Logistic Regression: 95% Naïve Bayes = 74% LSTM= 54–65%	-Labeling inconsistency.
[20] 2022	English	semantic Representat-ions	Machine Learning	BuzzFeed Political News	Random Forest 94.4% accuracy	-Dataset limited to political domain.
[28] 2024	English	semantic Representat-ions	Machine Learning , Deep Learning ,Transfer Features	Deepfake	logistic Regression + Transfer Features = 98.9% accuracy	- Domain specific dataset. - No real-time evaluation.
[29] 2023	Ukrainian/English	statistical + semantic Representations	Machine Learning	Kaggle Fake News dataset	High accuracy with Logistic Regression = 95 %	-lacks contextual features.
[30] 2023	English	statistical Representations	Ensemble ML	WELFake Dataset	LightGBM : Accuracy = 97.6%, F1 = 97.7%;	- Sensitive to textual noise, limited to text-only analysis.

[31] 2023	English	statistical Representations	Machine Learning	Kaggle Fake News dataset	Logistic Regression: Accuracy = 91%, Recall = 93.18%	- Feature Selection. - Deep Learning models.
[32] 2024	Arabic	statistical, semantic Representat-ions	Machine Learning	Arabic Sentiment Twitter Corpus	Accuracy = 89% (SVM with text + emoji features), F1= 0.88 to 0.89	-Some emoji meanings depend on context not captured numerically.

Table shows classic machine learning models (e.g., SVM, XGBoost) can achieve high accuracy over 95%, with the help of statistical features such as TF-IDF. However, their performance is still limited to specific domains (Single-domain) and lacks the contextual depth, which demonstrates the need for advanced models that can generalise and understand complex texts.

4.2 Sentiment With Deep Learning

Deep learning models have higher capability to grasp the emotion , contextual dependencies and subtle affective cues which older methods fail to capture . Thus, recent work has been devoted to the integration of sentiment and emotion variables into deep neural architectures for improved detection accuracy and durability .

To overcome this gap, recent work has explored augmenting traditional text representations with emotive and sentiment-based information. For instance, a complementary perspective on COVID-19 related fake-news tweets from a Kaggle dataset revealed that shallow statistical characteristics such as TF-IDF or Bag-of-Words paired with sequential deep learning architectures can achieve strong results. For instance, in this work, Bi-GRU achieved 91% accuracy, 93% recall and F1-score of 0.92, outperforming LSTM and other models [15]. This shows the importance of combining lexical cues with contextual sequence modeling. The Covid-19 Fake News Sentiment Analysis project extended this inquiry by engineering 39 features, such as sentiment, linguistic, and Named Entity Recognition (NER) features. Deep models on raw text only achieved 55-59% accuracy while AdaBoost fared somewhat better with 79.88%. However, the addition of these engineered features led to considerable improvements GRU, in particular, achieved 86.12% accuracy as well as improvements in AUC, precision, recall, and F1 score.[٢]

Another research extended this approach by combining textual, sentiment, and emotion data for further improvement of detection. Here, the semantic content of headlines was obtained using GloVe embeddings, the sentiment was obtained using the polarity scores of TextBlob, while the emotion features from NRCLex were used to identify fear, disgust, trust and joy categories, then aggregated into higher-level clusters such as novelty, expectation and neutrality. Only moderate performance was achieved by models trained only on text (LSTM, GRU, CNN, Bi-LSTM) but performance improved with constantly adding sentiment and emotion layers. The most striking result was achieved by Bi-LSTM with feature fusion, presenting AUC 96.77%, accuracy 96.89%, and F1 97.81%, showing proof that emotional cues, contained in the news content or reflected in the audience responses, give important signals in identifying false from real news.[١٧]

The study showed that combining emoji-based sentiment and textual sentiment considerably enhanced the performance. Inputs with merely text got an accuracy of roughly 73%, while adding emoji features raised the accuracy to 84%. When included in the proposed Bi-LSTM model, the overall accuracy achieved 99.68%, significantly exceeding both classic ML baselines and standard deep learning models. This highlights the importance of emojis as complementary sentiment signals to improve in fake news detection [12]. The profanity is not only verbal noise, but it contains emotional clues that boost the accuracy of

sentiment analysis. One LSTM model maintained the profanity while the other one removed it. The results indicated that removing the profanity decreased the accuracy from 83.4% to 81.6%. This finding implies that profane utterances include essential affective cues and they may lose subtle emotional information that negatively affects model performance when removed [33]. The above studies are concerned with incorporating sentiment and lexical data in false news detection, however the importance of advanced sentiment modeling has also been emphasized in comparable studies in other fields. For example, the OptiASAR system for healthcare reviews shows that the combination of BERT-based embeddings, aspect-level sentiment analysis, sequential models BiLSTM, GRU and attention mechanisms can significantly improve performance. Although such methods have not been used for fake news detection, they demonstrate the possibility of transferring aspect-based and domain-specific sentiment methods to better misinformation detection in the future [34]. Introduced by the study for Sentiment Analysis that directly addresses one of the fundamental shortcomings of traditional embedding models like Word2Vec and GloVe, i.e. their inability to capture emotional polarity. Traditional embeddings are good at capturing contextual and grammatical links, but they tend to place words with opposite sentiment, such as good and terrible, close together in the embedding space because they are used in similar contexts. To tackle this problem, the authors introduced the Continuous Sentiment Contextualized Vectors (CSCV) model that improves the pre-trained embeddings by adding the sentiment polarity obtained from lexicons such as SentiWordNet, SenticNet, and VADER. We adopt a modified CBOW architecture, which is trained to predict sentiment classes instead of words and then fuse the sentiment and semantic spaces using Principal Component Analysis (PCA) to get enhanced word representations that better capture emotional orientation. Extensive empirical evaluation on benchmark datasets MR, SST-1, SST-2 employing CNN, LSTM and Bi-LSTM architectures shows constant accuracy gains of 1-2% above state-of-the-art standard embeddings, with Bi-LSTM obtaining up to 88.6% accuracy. These results show that sentiment-aware embeddings can substantially improve the performance of downstream sentiment classification, offering more fine-grained affective modeling than typical semantic embeddings [35]. Deep learning models have made significant advances but performance still relies mainly on sequential architectures, which can face challenges in handling long-range relationships and subtle contextual inputs. These constraints motivate the study of more sophisticated language-understanding paradigms that can capture richer semantics, global context and cross-domain generalization. (Table 3) summarizes deep learning models with emotion features

Table 2. Deep learning models

REF/ Year	Language	Features represent-tation	Detection Method	Data Set	Performance	Limitations
[2] 2022	English	Statistical, semantic Representat-ions	Deep Learning + Machine Learning	WHO, Harvard Health, CDC, New York	GRU 86.12%, RNN 85.65%, with ~+20 AUC	-Source bias.
[15] 2024	English	Statistical Representa-tions	Deep Learning + Machine Learning	Kaggle	BiGRU: Accuracy = 0.91, Precision = 0.90, Recall = 0.93, F1 = 0.92. LSTM 0.90. ML models 0.88–0.89. CNN-1D Precision =	-Dataset limited -Did not address sarcasm, emojis, or hashtags. -Absence of explainability tools.

					1.00 Recall = 0.43	
[12] 2024	English	semantic Representat-ions	Deep Learning + Machine Learning	Tweets / X	- LSTM = 87% - Bi-LSTM model = 99.68%	-Sentiment limited to polarity. - Ethical aspects (bias, false positives, transparency) not addressed.
[17] 2023	English	semantic Representat-ions	Deep Learning + Machine Learning	Fakeddit	Bi-LSTM AUC = 96.77% Accuracy = 96.89% F1 = 97.81%	- limited for sarcasm, slang, multilingual. - Dataset imbalance.
[34] 2025	English	semantic Representat-ions	Deep Learning	Yelp	Accuracy =0.82, F1 = 75.5%	-Relies on labeled data.
[35] 2022	English	syntactic, semantic Representat-ions	Sentiment-enhanced word embeddings	IMDB, MR, SST-1, SST-2	Accuracy up to 88.6% (BiLSTM + refined Word2Vec)	- Limited to English language.

.

(Table3) illustrates the superiority of sequential deep learning models (such as Bi-LSTM and BiGRU) in achieving exceptional accuracy, approaching 99%, when combining semantic and statistical features. Despite this high performance, these models share fundamental challenges, most notably their inability to understand linguistic complexities (such as sarcasm and colloquialisms) and their lack of transparency and explainability.

4.3 Transformers Models

Traditional word representations have showed a good ability to capture surface-level linguistic patterns, particularly when paired with deep learning methods like convolutional and recurrent neural networks. Their biggest disadvantage nonetheless is the lack of a representation of meaning in an integrated, context rich framework. Modern models, on the other hand, leveraging transformer topologies, have demonstrated improved performance owing to their capability to encode language in deeper semantic spaces and to capture subtle differences in style, tone and contextual dependencies. Following recent discoveries, scholars also started to distinguish between human-made and AI-made fake news. Human-made fakes usually depend on rhetorical devices and cultural signs, while AI-made texts exhibit

particular stylistic signs that can be identified by using contextual embeddings and transformer-based representations .[٣٦]

The current study developed a new dataset of 1500 Modern Standard Arabic articles 500 real pieces obtained from Okaz newspaper, 500 fake articles produced by human writers, and 500 fake articles generated using language models (GPT). Experiments included classic word representations Word2Vec, FastText and GloVe, and contextual representations based on transformer models such as ARBERT,

MARBERT, and AraVec. Different deep architectures such as CNN, Bi-LSTM, and GRU were also used along with transformer based models to differentiate real and fraudulent news. The results showed that the human performance for fake news detection was limited, not exceeding 52% for AI-generated texts. However, the transformer models, ARBERT in particular, produced much better results, with an overall accuracy of about 78%, outperforming traditional models and human participants [37]. Meanwhile, this line of research was continued by building a cross-domain dataset across politics, economy, health, religion and manufactured stories characterized by risk level high vs. low. The model used MARBERTv2, QARiB embeddings, and a multitask deep learning framework to do three tasks simultaneously: fake-news detection, categorical categorization, and risk prediction. The system attained F1-scores of 94.12%, 84.92% and 88.91% for detection, categorization and risk assessment respectively. Notably, ArabFake also suggested valence scoring as an explainability technique [38]. In the same line, a model was proposed to push the field of sentiment analysis further with the concept of an integrated deep architecture, called Deep Sentiment Analysis (DSA), that depends on a Decision-Based Recurrent Neural Network (D-RNN). The proposed model uses the capability of BERT-Large Cased to deeply comprehend the linguistic context and the Stochastic Gradient Descent (SGD) algorithm to improve the training process. Besides, it employs feature extraction methods such as Bag-of-Words and Word2Vec to represent the text more precisely. Furthermore, the model achieves better performance by combining Aspect-Based Sentiment Analysis (ABSA) and Priority-Based Sentiment Analysis (PBSA) for more complete and accurate sentiment categorization. The experimental findings indicated that the suggested model had a better accuracy of sentiment prediction than existing models [39]. The LGCF model is a complementary framework to the DSA framework with the same goal of modeling better understanding of emotions via integrating local and global context analysis. It proposes a multilingual architecture that employs attention mechanisms and bidirectional networks to produce accurate representations of semantic and affective interactions in text. Thus, LGCF can be considered a natural extension of the DSA approach, improving the ability of models to capture complex emotional characteristics and offering possibilities for a wider use of fake news detection based on the study of contextual and affective tone.[٤٠]

These results show that more sophisticated contextual models, especially Transformers, are more robust and reliable for fake-news identification than traditional techniques. As misinformation gets more sophisticated, especially with the advent of AI-generated content, future systems will need to leverage deep contextual knowledge and transformer-based architectures in order to stay effective. (Table 4) summarizes the transformer models.

Table 3. transformer models

REF/ Year	Language	Features represent-tation	Detection Method	Data Set	Performance	Limitations
[37] 2025	Arabic	Contextual, semantic Representat-ions	Deep Learning and Transform er-based models	Real (Okaz), Human-written Fake, GPT-generated Fake	-Humans: 77.8% (real), 48.2% (human fake), 52.7% (GPT fake) - DL models: max =73% (human fake) and =62% (GPT fake) - Transforme	- GPT-generated articles highly similar to human-written ones.

					rs: up to 92% (human fake) and 78% overall with ARBERT	
[38] 2024	Arabic	Contextual, semantic Representat-ions	Multitask Deep Learning Framework using MARBERTv2	ArabFake Dataset	Fake news detection: F1 = 94.12% Categorization: F1 = 84.92% Risk prediction: F1 = 88.91%	-Imbalanced dataset.
[39] ٢٠٢٣	English	Contextual and Semantic Representations (BoW, Word2Vec, BERT-large-cased)	Deep Sentiment Analysis (DSA) using Decision-based Recurrent Neural Network (D-RNN) combined with ABSA and PBSA	Laptop-ACOS, Restaurant, Twitter (from Kaggle)	Laptop: Accuracy = 95.9% Restaurant: Accuracy = 85.31% Twitter: Accuracy = 86.0%	- Requires high hardware configuration (e.g., 32 GB GPU) for training . - Limited to text only; cannot currently extract sentiments from images or videos.
[40] 2022	Multilingual (English & Chinese)	Local and Global Context representations (BERT embeddings), Context-features Dynamic Mask (CDM) and Weighting (CDW)	Aspect-Based Sentiment Analysis (ABSA) utilizing LGCF model (BGRU + CNN + MHSA)	3 Chinese datasets (Camera, Phone, Car) & 6 English datasets (Laptop14, Restaurant14/16, Twitter, Tshirt, Television)	Accuracy up to 98.26% on Car dataset and 91.87% on Twitter dataset	- Sentences with neutral sentiment polarity are more prone to misclassification. - Integrating excessively rich global context can introduce noise and overfitting.

Table (4) shows that Transformer models such as BERT and its derivatives (ARBERT and MARBERT) achieve a remarkable breakthrough in the accuracy of fake news detection and sentiment analysis, averaging over 94% across multiple languages. Despite this superiority, these models remain limited by their high hardware costs and their sensitivity to noise in neutral texts.

4.4 Model Optimization Strategies In Fake News Detection

To improve the resilience and generality of fake-news detection systems, several fundamental issues need be addressed, including data distribution, model complexity, and feature representation challenges. One of the most persistent problems is class imbalance , which often results in models being biased towards the majority class , usually actual news , while reducing their capacity to correctly identify fake news . This imbalance increases the probability of type II errors of not detecting bogus content and lowers the reliability of detection systems in the fight against disinformation. To address this, the designers of Themis model applied data augmentation techniques like TSIT and MixGen to rebalance the dataset and increase the training samples. The experimental results indicated that the total performance was significantly improved, with the ReCOVery dataset obtaining up to 97.5% accuracy by combining TSIT with LoRA, suggesting that augmentation-based rebalancing offers a feasible and viable way to address data imbalance [26]. Other than balancing at the data level, a crucial strategy to improve performance is to optimize model efficiency. The fine-tuning of big language models such as BERT suffers from severe computing hurdles because of the huge number of parameters, which typically limits their implementation in large-scale or resource-constrained scenarios. In this respect, the ABERT model provided a parameter-efficient alternative by freezing the base weights of BERT and solely training light-weight adapters. This strategy decreased the number of trainable parameters by ~67.7% compared to full fine-tuning and achieved great accuracy (91.98% on PolitiFact, 85.79% on GossipCop and 99.09% on an AI-generated news dataset). These results demonstrate the feasibility of ABERT to obtain competitive results against heavier systems like BERT and DistilBERT, while significantly reducing the computational costs.[٤١]

Furthermore, the influence of input encoding schemes on the performance optimization of fake-news detection has been pointed out in recent works. In a study to evaluate the SGDM-GRU model, three encoding methodologies profile based, Word2Vec and BERT were tested to determine their relative impact on classification accuracy . Profile based encoding resulted in the lowest results, 88.7% accuracy on Weibo and 89.7% on Twitter, whereas Word2Vec produced slightly higher accuracy, 89.7% and 89.4% respectively. However, with BERT encoding, the model achieved an impressive accuracy of 98.3% on Twitter, with an F1-score and recall of 98.5% [22]. These results provide evidence that contextualized embeddings such as BERT encode deeper semantic and syntactic relationships that standard encodings miss. This, in turn, dramatically increases the accuracy and the generalization of fake-news detection systems .

In general, these optimization methodologies, from data balance to parameter efficient fine tuning and sophisticated input encoding, show that enhancing fake-news detection needs an integrated approach that addresses both data and model level issues.

4.5 Graph-Based Models

A challenge for heterogeneous graph-based fake news detection systems is the poor representation of short texts. Most of the news shared on social media platforms are usually short and concise, which results in less semantic information and difficulty to extract discriminative features that allow the model to tell the difference between true and false news. Existing models heavily depend on linking entities to external knowledge bases, but they are usually not effective for short texts and different writing styles, which limits the ability of models to capture accurate semantic relations. Although these methods have facilitated knowledge-based detection, they largely ignore the topic context of the news and heavily rely on the quality of entity knowledge alignment, which makes them vulnerable to semantic ambiguity and polysemy errors. Constructing. The researchers used the LUN and SLN datasets which have multiple news types like trusted, satire, hoax, and propaganda for evaluating the model in both binary and four-way classification settings. The data were modeled as a heterogeneous graph of three node types, sentences, topics and entities, where topics were extracted via Latent Dirichlet Allocation (LDA) and entities were linked to Wikipedia using TAGME. The proposed FND model incorporated a two-layer heterogeneous

graph attention mechanism to fuse textual, knowledge-based and topical information in a unified semantic framework. The experimental results have shown that the integration of the topic information can significantly improve the performance of baseline models, including BERT and CompareNet, increasing the accuracy of binary classification and four-way classification by 0.83% and 1.85%, respectively, thus confirming the effectiveness of the topic integration in fake news detection [42]. A study [43] aimed to enhance the accuracy of automated rumor verification in real-world scenarios, relying on the analysis of Twitter claims as events unfolded and developed using the PHEME database. To achieve this, researchers devised a time-bound mechanism for collecting external evidence to ensure the exclusion of any information published after the rumor's emergence and to avoid knowledge contamination by future events. They also developed an innovative approach to constructing knowledge graphs capable of identifying paths and connections between disparate pieces of information. Furthermore, they created a novel sequence-matching metric to select the most relevant statements to the rumor, surpassing traditional information retrieval methods. These statements were then integrated as text with the original tweet into the verification model. Finally, they refined and updated the PHEME database classifications, specifically the "Unverified" category, after discovering discrepancies and inconsistencies between the retrieved evidence and the original classifications. The results proved that this proposed model outperformed previous state-of-the-art models in performance on the PHEME base, in addition to showing exceptional ability and high flexibility in generalization when tested on time-separated and objectively different data, specifically on a database related to Covid-19 pandemic rumors (Covid-RV). To address the limitations of existing multi-task learning models in processing long conversation threads and capturing inter-task dependencies, a study [44] proposed the Coupled Hierarchical Transformer for stance-aware rumor verification. This novel architecture adapts pre-trained contextualized embeddings, specifically BERT, by dividing lengthy social media threads into shorter subthreads to capture local interactions, which are then sequentially processed by a global Transformer layer to encode the overall conversation context. To further exploit the relationship between stance classification and rumor verification, the researchers introduced a coupled transformer module that explicitly models inter-task interactions, alongside a post-level attention mechanism that integrates predicted stance labels directly into the rumor verification process. Extensive evaluations on the SemEval-2017 and PHEME benchmark datasets demonstrated the model's superiority, as it significantly outperformed state-of-the-art multi-task learning baselines in terms of Macro-F1 scores. a study [45] introduced an innovative framework for rumor verification on social media platforms, relying on Stance-Aware Structural Modeling to overcome the sequence length constraints that limit the efficiency of Large Language Models (LLMs) when processing long conversations. The proposed methodology entails encoding each post and integrating it with its specific stance signal, followed by aggregating and compressing reply embeddings based on stance categories to form a dense, semantically enriched representation of the entire conversation thread. To enhance the structural understanding of the context, the researchers incorporated two key structural covariates: "stance distribution" to capture collective orientation and stance imbalance and "hierarchical depth level" to measure the influence and progression of replies within the conversation. Extensive experiments on benchmark datasets demonstrated this model's significant superiority over previous approaches in predictive accuracy, alongside proving its high effectiveness in early rumor detection and its seamless cross-platform generalization capabilities. (Table 5) summarizes the graphical models used in detecting fake news.

Table 4. Graph-Based Models

REF / Year	Language	Features represent-tation	Detection Method	Data Set	Performance	Limitations
[42] 2024	English	Semantic, Statistical Representa-tion	Heterogeneous Graph Attention Network	LUN SLN	Accuracy = 96.5% (2-way) Accuracy = 95.7% (4-way)	- Performance depends on entity linking quality.
[43] ٢٠٢٤	English	Text integrated with external evidence, Knowledge Graphs (identifying paths and connections), and a novel sequence-matching metric	Automated rumor verification model combining time-bound external evidence retrieval and Knowledge Graph construction	PHEME dataset and Covid-RV dataset	Accuracy = 0.523	Requires strict time-bound mechanisms to collect evidence to avoid knowledge contamination by future events.
[44] 2022	English	embeddings (BERT), Hierarchical Transformer (local subthreads + global conversation context encoding), and predicted stance labels	BranchLSTM + NileTMRG. MTL2 (Veracity+Stance).	SemEval-2017 and PHEME benchmark datasets	Accuracy of pheme= 0.466 SemEval=0,678	Significantly outperformed state-of-the-art multi-task learning baselines in terms of Macro-F1 scores

Table (5) shows that graph-based models and hierarchical transformers excel at integrating external knowledge and analyzing the contextual structure of conversations. Despite their remarkable superiority, their effectiveness is heavily dependent on the quality of entity linking and requires rigorous time-based evidence-gathering mechanisms.

4.6 Hybrid Deep Learning

The ongoing challenge of fake news detection stems from the inherent intricacies of misinformation itself, including its capacity to mix factual fragments with fabricated ones, employ various linguistic styles, and utilize textual and visual modalities to deceive audiences. Traditional deep learning and machine learning models, successful as they are at classifying binary cases of true vs false news, often fail to generalize across domains, or capture subtle interplay between semantics, context and modality. This analytical limitation has motivated researchers to search for models that can capture not only the linguistic structure of news content, but also its contextual and cross-modal relationships. To tackle this challenge, recent

studies have proposed hybrid deep learning models by combining different neural architectures or multiple feature types to improve the representational depth. Some recent advances on the detection accuracy and robustness have been made by the methods of hybridization of CNN and BiLSTM, fusion of graph-based propagation models with temporal networks, and integration of textual and visual encoders. In addition, we have stressed the necessity of multi-class classification, enabling us to classify misinformation into more specific classes such as mainly true or mostly false, reflecting the intricacies of real-world scenarios. Still, there are conceptual and methodological limits in the field. Many hybrid systems improve accuracy but suffer from interpretability and scalability issues. Higher architectural complexity generally leads to lower transparency and computing efficiency. Similarly, multi-class situations are under-explored and appraised inconsistently across datasets, making the results hardly comparable. This calls for a detailed review of the latest hybrid frameworks to examine the role of alternative architectural combinations in addressing the limitations of traditional models and extending the frontiers of false news detection research, whether textual, multimodal, or reasoning-based. Methods like Bag-of-Words or TF-IDF, however, do not take into account the semantic context or deeper linguistic patterns. To overcome this limitation, the researchers propose the CNN-3BiLSTM model that combines the capability of Convolutional Neural Networks (CNN) to extract local patterns and key phrases and the power of Bidirectional Long Short-Term Memory (BiLSTM) networks to capture long-range contextual dependencies within the text. The experiments on several datasets such as Fake News Corpus, Kaggle Fake News, GossipCop and Fakeddit show that the proposed model outperforms traditional baselines by a significant margin, with the classification accuracy being improved by around 8% to 15%, especially on short text datasets like GossipCop. These results validate the efficacy of merging CNN with BiLSTM [46]. Recently, hybrid multi-feature models have drawn more and more attention to capture varied textual representations and contextual dependency. A prominent contribution in this area is a study that provides an integrated framework of TF-IDF networks, convolutional neural networks (CNNs), and bidirectional long

Short term memory networks (BiLSTM) followed by a fast learning network (FLN) classifier. The model combines the global, spatial, and temporal aspects at an early stage, thus, exploiting the statistical and semantic qualities of news material. We experimentally evaluated our model on benchmark datasets ISOT and FA-KES and it outperformed current models such as HyproBERT and DeepCnnBilstm with over 99% accuracy. The results of this technique show the efficacy of hybrid deep learning architectures combining several feature domains and providing high detection accuracy along with the computational efficiency achieved by the use of FLN as a lightweight and fast classifier.[٤٧]

In addition to sequential and convolutional combinations, hybridization at graph and temporal levels has also been examined by researchers to capture the structural dissemination of news in social platforms. The study included four important datasets in the field of false news identification, including Weibo, Twitter (twitter15, twitter16), Politifact and Gossipcop, covering a variety of genres ranging from political rumors to entertainment news. The researchers used a hybrid model that incorporates Spectral Graph Deep Learning (SGDM) that captures the structural patterns of news transmission, and a GRU network that learns the temporal and dynamic elements. The model was compared with various baseline methods including Decision Tree, SVM, GCN and LSTM. The results showed that the SGDM-GRU greatly outperformed the others, reaching the best performance in all datasets. On Twitter, it reached an accuracy of 98.3% with a recall of 98.5% and an F-score of 98.5%. It also achieved excellent results on the other datasets (Weibo: 0.9205, Politifact: 0.9732, Gossipcop: 0.9642), demonstrating its capability to detect bogus news with high precision [22]. The sentiment analysis role is a strong indicator of the manipulative content. A study on 402 tweets using XGBoost found that the sentiment was the most influential factor, followed by the network centrality, word similarity, and the interaction volume. The Random Forest algorithm had the highest accuracy (94.1%) among the models.[١٩]

Building on such hybrid concepts, models have begun to include domain adaptation techniques to better generalization over diverse forms of news information. The SLFEND (Soft-Label for Multi-Domain Fake News Detection) model has a hybridization type that blends various neural architectures into one textual framework. The hybridization combines Leap-GRU that improves efficiency by leaving out uninformative terms, and a group of expert networks (TextCNN, DCNN and DPCNN) that are coordinated by a Domain Gate to balance their domain-specific contributions. This hybrid technique attempts to harness the

complimentary power of recurrent and convolutional models, modeling both sequential dependencies and local textual patterns, while alleviating domain bias. As a result, SLFEND delivers better generalization on many domains and better detection accuracy than standard single-model baselines. Results revealed better performance with F1 scores of 92.49% on Weibo21 and 89.98% on Thu, outperforming state-of-the-art baselines [48]. Further, hybrid deep learning has developed with intricate textual architectures, incorporating linguistic, relational, and stance-based elements. For a more sophisticated approach, the Joint Learning Framework for Fake News Detection (JLFND) enhances the power of deep learning by adding entity recognition (NER), relational features (RFC) and stance detection (SD) to BERT. This paradigm incorporates many signals into a coherent model with hierarchical attention and consistency losses. JLFND outperformed both traditional baselines and typical BERT implementations with impressive accuracy F1 of 0.94/0.95 and 0.98/0.98 on the FakeNewsNet datasets PolitiFact and GossipCop respectively.[٧]

Another area for hybrid model development has been the use of Chain-of-Thought reasoning to boost the interpretability and reliability of false news detection systems. This method enables the model to demonstrate the logical reasoning stages to get its decision, so that the user may understand why a piece of news is labeled as true or false. Also, it helps mitigate the impact of random errors by allowing the model to revisit the reasoning sequence and check the internal coherence before the final result is generated, which makes the classification process more accurate and transparent. Within this framework, a hybrid model combining Chain-of-Thought reasoning and In-Context Learning with Knowledge Distillation from the GPT-4 model to the lighter Qwen2 model was built. The study used the CEFKE dataset, which was created to facilitate linguistic and logical interpretability in classification judgments. The CEFKE dataset added a category of "undetermined" news, which is a signal of a progression towards multi-level categorization. The findings show the better performance of the model with respect to the accuracy (94%) and the interpretative consistency .[٤٩]

The methodological path has also gone beyond text-only designs to multimodal fusion, combining textual and visual clues for a better detection of manipulation. The MultiBan FakeDetect study goes beyond text-based methods and demonstrates the potential of multimodal solutions. The researchers worked with a dataset of 9,600 text-image combinations, across four categories including disinformation, rumor, clickbait and true news. They integrated language models such as mBERT and XLM-RoBERTa with the visual model DenseNet-169, investigating three fusion strategies Early, Late and Intermediate. The early fusion approach among them performed best with an accuracy of 79.69% for binary classification, whereas that for multi-class classification was more difficult due to the overlap among categories .[٢٤]

This line of research led to more advanced multimodal architectures that focused on semantic alignment between text and images. For example, the SSA-MFND model was assessed on datasets in Chinese Weibo and in English. It extracted textual characteristics using BERT with BERT Whitening to address embedding anisotropy and visual features using VGG-19. with the addition of OCR for text in photos and tools for entity recognition and description. By using a Co-Attention strategy to maintain the semantic consistency at both the entity and description levels, SSA-MFND achieved state-of-the-art results of 0.912 accuracy and 0.914 F1 on Weibo and 0.975 accuracy and 0.973 F1 on English datasets, respectively. We show that the combination of textual and visual modalities coupled with embedding refinement leads to better and more generalizable outcomes compared to text-only techniques.[٨]

This approach was subsequently enhanced in subsequent investigations by explicitly modeling cross-modal alignment between written text and related imagery, and handling circumstances of semantic incongruence. The drawback of unimodal models in detecting false news, stating that text-based methods usually neglect the discrepancy between the text content and the associated graphics, letting misleading stories slip through if real text is combined with altered visuals. To address this problem, the authors proposed the MFFFND-Co model that integrates textual features extracted from BERT, spatial features of images using VGG19, and frequency-domain features using DCT-CNN. They also added a semantic consistency module to evaluate the consistency between the news text and the automatically generated image captions. The experimental findings reveal the reasonably good performance of textual characteristics only, with the classification accuracy beyond 85%. But, the results were greatly improved with the addition of picture features, with the F1-score of 90% for the Weibo dataset and 94% for the English News dataset. The comparison demonstrates that images, especially when their frequency data are

extracted and merged with text using the Co-Attention mechanism, boost the performance of the model for fake news detection, particularly when the textual and visual modalities are semantically contradictory [25]. The researchers used a multimodal approach to detect fake news. This included natural language processing using LSTM to analyze text, computer vision with ANFIS to detect picture modification, and fuzzy matching to evaluate the consistency between text and image. The outputs were then integrated into a fuzzy inference system, yielding an accurate and interpretable conclusion.[٢٧]

In sum, these results point to a clear trend in the field: the development in false news detection is more and more being achieved via hybrid and integrative methods. Researchers are shifting from statistical and lexical characteristics to multimodal fusion and context-rich architectures, producing systems that are more accurate, resilient, and adaptable to other domains.)Table 6(provides a summary of the hysterical patterns in fake news detection.

Table 5. Hybrid deep learning

REF/ Year	Language	Features represent-tation	Detection Method	Data Set	Performance	Limitations
[7] 2025	English	semantic consistency Representat-ions	Enhanced BERT with Multi-task Learning	FakeNewsNet	PolitiFact: Acc = 0.94, F1 = 0.95 GossipCop: Acc = 0.98, F1 = 0.98	-High computational cost.
[19] 2023	English	Contextual, Statistical, Semantic Representations	Machine Learning	402 tweets from Twitter (X)	Random Forest with accuracy = 94.1%	- Very small data size. - Does not take into account cultural and linguistic diversity.
[22] 2025	English	Semantic Representations	Spectral Graph Deep Learning Model (SGDM) + Gated Recurrent Unit (GRU)	Weibo, X, Politifact, Gossipcop	Accuracy up 98.3 with X	- Performance sensitive to missing/imbalanced features.
[8] 2022	Chinese English	semantic consistency Representat-ions	Hybrid multimodal model	Weibo dataset (Chinese), English multimodal dataset	Weibo Accuracy = 0.912, F1 = 0.914; English: Accuracy = 0.975, F1 = 0.973	-Reliance on external tools may introduce errors/bias.

[27] 2026	English	statistical	Hybrid multimodal model	X platform, Buzzfeed, PolitiFact	Accuracy of X= 0.941 , PolitiFact = 0.903, Buzzfeed = 0.893	-higher complexity.
[25] 2025	Chines English	Contextual, semantic Representat-ions	Hybrid multimodal model	Weibo & English News	F1-score = 90.0%	- High computa-tional cost.
[24] 2025	Bangla	Contextual, semantic Representat-ions	Hybrid multimodal model	MultiBanFakeDetect Dataset	Accuracy = 79.69% (Early Fusion)	-Limited discussion of ethical and practical.
[46] 2023	English	Contextual, semantic Representat-ions, Network-based	Deep Learning	Fake News Corpus, Kaggle Fake News, GossipCop, Fakeddit, Twitter15	accuracy of CNN-3BiLSTM 92 to 93%	-propagation model assumes complete info; real-time scalability challenges.
[48] 2023	Chinese	Contextual, semantic Representat-ions, Soft Labels	Hybrid model	Weibo21, Thu	Weibo21: F1 = 92.49% Thu: F1 = 89.98%	- Training. time is not the shortest compared to some baselines.
[49] 2025	Chinese	semantic Representat-ions	Hybrid method	CEFKE Dataset	F1: 93.28%	- Retrieval-based learning adds computational cost. - Undeterminedlabel may be overused in ambiguous cases.

Table (6) shows that hybrid and multimodal models achieve exceptional accuracy, often exceeding 94%, thanks to the integration of textual, visual, and network features. However, this high accuracy comes at the cost of significant structural complexity, high computational costs, and scalability challenges.

5.☐ Major Limitations And Future Research Directions In Fake News Detection Models

The above study raises a number of technical and methodological issues. Future work needs to address these issues to further improve false news detecting systems.

☐☐ Sarcasm and irony: The most challenging problems in preprocessing as mentioned in research [12], [5], [15]. There is no adequate paradigm to handle sarcasm and irony.

Slang: Slang was considered as a challenge during the preprocessing step in the research [21 .[١٤] ,[

Multilingual data: This was highlighted as a barrier in preprocessing [6], [24] and the study [28] was limited to only processing two languages, Chinese and English.

- Multilingualism: The propagation of news is one of the concerns most explored in studies on fake news detection. It is a frequent problem in all communities. Designing a system that is only capable of processing one language limits the generalizability of the model. Systems that work in more than one language are needed.[٤]
- Multimedia content: The growth of multimedia on social media is on the rise. Fake news is commonly accompanied by photographs or video, hence it is necessary to include multimedia in news identification. A model that mixes text and image/video needs to be designed [7.[١٨] ,[
- Emojis and Paralinguistics : Emojis are used to represent user emotions such as, rage, excitement, anxiety, and other emotions . Their removal from the input set has an adverse effect on the quality of the sentiment analysis. Therefore, emojis are better to be used as features to feed into the system, instead of eliminating them .[٢٨]
- Labeled data dependency: Over-dependence on labeled data hinders generalization and propagates errors from the preceding step. Semi-supervised or unsupervised learning is more acceptable [12 .[٢١] ,[
- Real time detection: Detecting bogus news before it is circulated might limit its detrimental impact. Some of the proposed systems require non real-time signals to detect news, e.g. based on its speed of spreading across the network. Therefore, it is necessary to develop a strategy that evaluates in real time .[١٦]

- Coverage and cross-platform generalization limitations. Using narrow or specialized datasets, e.g. Reddit [17], Kaggle [6], limits coverage and generalization.
- Class imbalance: Fake news datasets are usually imbalanced with one class resulting in bias and poor generalization ABERT found this problem and offered data augmentation and rebalancing as future remedies.[١٧]
- Transparency and interpretability: To avoid uncertainty and undermine the recipient's trust in classifying a piece of news as false, complex models should not operate as a "black box" as in studies [17]. Study [27] directly addressed this issue by incorporating an interpretability and transparency mechanism Therefore, by using interpretability mechanisms such as SHAP or LIME users can understand the rationale behind classification decisions.

6. Conclusion

This review demonstrates that fake news detection transcends simple linguistic analysis toward a comprehensive understanding that integrates semantics, emotions, and social context. While traditional models achieve acceptable performance in limited domains, deep, transformers and hybrid particularly multimodal models exhibit greater generalization ability and higher accuracy in real-world settings.

Despite these advances, challenges such as training cost, model interpretability, real-time detection, AI-generated content, linguistic bias remain and demand systematic solutions. The study therefore recommends a future focus on developing interpretable hybrid models that combine the predictive power of transformers with transparency in reasoning, while incorporating ethical and linguistic considerations to ensure the accuracy and efficiency of fake news detection systems across languages and cultures.

Acknowledgements: The authors would like to thank Mohammed Al-khafajiy for their valuable support and contributions to this work.

Funding Information: This research received no external funding.

Authors Contributions: Zainab Khaled Abdullah: Conceptualization, methodology, data curation, writing – original draft. Ahmed Jassem Mohammed: Supervision, validation. Both authors read and approved the final manuscript.

Conflicts of Interests: The authors declare that they have no conflicts of interest.

Ethical Approval: This study did not involve human or animal subjects, and ethical approval was not required.

Data Availability Statements: The data supporting the findings of this study are available within the article and its references. No new underlying datasets were generated during this systematic literature review.

Reference :

[١] Pew Research Cente, “Social Media and News Fact Sheet,” Pew Research Center’s Journalism Project. Accessed: Sep. 23, 2025. [Online]. Available: https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/

[٢] C. Iwendi, S. Mohan, S. khan, E. Ibeke, A. Ahmadian, and T. Ciano, “Covid-19 fake news sentiment analysis,” Computers and Electrical Engineering, vol. 101, Jul. 2022, doi: 10.1016/j.compeleceng.2022.107967.

[٣] Y. Mendes Rocha et al., “The impact of fake news on social media and its influence on health during the COVID-19 pandemic: a systematic review”, doi: 10.1007/s10389-021-01658.

[٤] F. N. Selnes, “Fake news on social media: Understanding teens’ (Dis)engagement with news,” Media Cult. Soc., vol. 46, no. 2, pp. 376–392, Mar. 2024, doi: 10.1177/01634437231198447.

[٥] E. T. Zanatta, G. P. De Macedo Wanderley, I. K. Branco, D. Pereira, L. H. Kato, and E. M. C. P. Maluf, “Fake news: The impact of the internet on population health,” Rev. Assoc. Med. Bras., vol. 67, no. 7, pp. 926–930, 2021, doi: 10.1590/1806-9282.20201151.

[٦] M. A. B. Al-Tarawneh, O. Al-irr, K. S. Al-Maaitah, H. Kanj, and W. H. F. Aly, “Enhancing Fake News Detection with Word Embedding: A Machine Learning and Deep Learning Approach,” Computers, vol. 13, no. 9, Sep. 2024, doi: 10.3390/computers13090239.

[٧] M. Abdullah, Z. Hongying, A. Javed, O. Mamyrbayev, F. Caraffini, and H. Eshkiki, “A joint learning framework for fake news detection,” Displays, vol. 90, Dec. 2025, doi: 10.1016/j.displa.2025.103154.

[٨] W. Shang, K. Song, J. Ji, T. Yi, J. Cai, and X. Li, “Semantic space aligned multimodal fake news detection,” Information Fusion, vol. 125, Jan. 2026, doi: 10.1016/j.inffus.2025.103469.

[٩] S. Kuntur, A. Wróblewska, M. Paprzycki, and M. Ganzha, “Under the Influence: A Survey of Large Language Models in Fake News Detection,” IEEE Transactions on Artificial Intelligence, vol. 6, no. 2, pp. 458–476, Feb. 2025, doi: 10.1109/TAI.2024.3471735.

[١٠] A. Sandu, I. Ioanăș, C. Delcea, M. S. Florescu, and L. A. Cotfas, “Numbers Do Not Lie: A Bibliometric Examination of Machine Learning Techniques in Fake News Research,” Algorithms, vol. 17, no. 2, Feb. 2024, doi: 10.3390/a17020070.

[١١] M. A. Alonso, D. Vilares, C. Gómez-Rodríguez, J. Vilares, and G. Lys, “electronics Sentiment Analysis for Fake News Detection,” 2021, doi: 10.3390/electronics.

[١٢] A. Bhardwaj, S. Bharany, and S. K. Kim, “Fake social media news and distorted campaign detection framework using sentiment analysis & machine learning,” Heliyon, vol. 10, no. 16, Aug. 2024, doi: 10.1016/j.heliyon.2024.e36049.

[١٣] K. Shu, S. Wang, and H. Liu, “Beyond news contents: The role of social context for fake news detection,” in WSDM 2019 - Proceedings of the 12th ACM International Conference on Web Search and Data Mining, Association for Computing Machinery, Inc, Jan. 2019, pp. 312–320. doi: 10.1145/3289600.3290994.

[١٤] M. Sudhakar and K. P. Kaliyamurthie, “Detection of fake news from social media using support vector machine learning algorithms,” Measurement: Sensors, vol. 32, p. 101028, Apr. 2024, doi: 10.1016/j.measen.2024.101028.

[١٥] M. T. Zamir, F. Ullah, R. Tariq, W. H. Bangyal, M. Arif, and A. Gelbukh, “Machine and deep learning algorithms for sentiment analysis during COVID-19: A vision to create fake news resistant society,” PLoS One, vol. 19, no. 12 December, Dec. 2024, doi: 10.1371/journal.pone.0315407.

[١٦] D. G. Dev, V. Bhatnagar, B. S. Bhati, M. Gupta, and A. Nanthaamornphong, “LSTMCNN: A hybrid machine learning model to unmask fake news,” Heliyon, vol. 10, no. 3, Feb. 2024, doi: 10.1016/j.heliyon.2024.e25244.

[١٧] S. K. Hamed, M. J. Ab Aziz, and M. R. Yaakub, “Fake News Detection Model on Social Media by Leveraging Sentiment Analysis of News Content and Emotion Analysis of Users’ Comments,” Sensors, vol. 23, no. 4, Feb. 2023, doi: 10.3390/s23041748.

[١٨] K. Atchariyachanvanich, C. Saengkhunthod, P. Kerdnoonwong, H. Chanlekha, and N. Cooharojananone, “Improvement of a Machine Learning Model Using a Sentiment Analysis Algorithm to Detect Fake News: A Case Study of Health and Medical Articles on Thai Language Websites,” Journal of Cases on Information Technology, vol. 26, no. 1, 2024, doi: 10.4018/JCIT.344812.

[١٩] M. Park and S. Chai, “Constructing a User-Centered Fake News Detection Model by Using Classification Algorithms in Machine Learning Techniques,” IEEE Access, vol. 11, pp. 71517–71527, 2023, doi: 10.1109/ACCESS.2023.3294613.

[٢٠] N. Seddari, A. Derhab, M. Belaoued, W. Halboob, J. Al-Muhtadi, and A. Bouras, “A Hybrid Linguistic and Knowledge-Based Analysis Approach for Fake News Detection on Social Media,” IEEE Access, vol. 10, pp. 62097–62109, 2022, doi: 10.1109/ACCESS.2022.3181184.

[٢١] V. H. Nguyen, K. Sugiyama, P. Nakov, and M. Y. Kan, “FANG: Leveraging Social Context for Fake News Detection Using Graph Representation,” Commun. ACM, vol. 65, no. 4, pp. 124–132, Mar. 2022, doi: 10.1145/3517214.

[٢٢] A. Sahi et al., “SGDM-GRU: Spectral graph deep learning based Gated Recurrent Unit model for accurate fake news detection,” Expert Syst. Appl., vol. 281, Jul. 2025, doi: 10.1016/j.eswa.2025.127572.

[٢٣] Q. Liu, Y. Lyu, J. Tang, and W. Fan, “Optimizing the Service Efficacy of Crowd Ratings in Curbing Fake News Dissemination on Social Media,” International Journal of Crowd Science, vol. 8, no. 3, pp. 110–121, Sep. 2024, doi: 10.26599/IJCS.2024.9100020.

[٢٤] F. T. J. Faria et al., “MultiBanFakeDetect: Integrating advanced fusion techniques for multimodal detection of Bangla fake news in under-resourced contexts,” International Journal of Information Management Data Insights, vol. 5, no. 2, Dec. 2025, doi: 10.1016/j.jjimei.2025.100347.

[٢٥] J. Cao et al., “Fake News Detection Based on Cross-Modal Ambiguity Computation and Multi-Scale Feature Fusion,” Computers, Materials and Continua, vol. 83, no. 2, pp. 2659–2675, 2025, doi: 10.32604/cmc.2025.060025.

[٢٦] D. A. Mura et al., “Is it fake or not? A comprehensive approach for multimodal fake news detection,” Online Soc. Netw. Media, vol. 47, Jul. 2025, doi: 10.1016/j.osnem.2025.100314.

[٢٧] T. M. H. Gedara, V. Loia, and S. Tomasiello, “A fuzzy-based multimodal approach for interpretable fake news detection,” Appl. Soft Comput., vol. 179, Jul. 2025, doi: 10.1016/j.asoc.2025.113277.

[٢٨] M. Khalid et al., “Novel Sentiment Majority Voting Classifier and Transfer Learning-Based Feature Engineering for Sentiment Analysis of Deepfake Tweets,” IEEE Access, vol. 12, pp. 67117–67129, 2024, doi: 10.1109/ACCESS.2024.3398582.

[٢٩] A. Mykytiuk, V. Vysotska, O. Markiv, L. Chyrun, and Y. Pelekh, "Technology of Fake News Recognition Based on Machine Learning Methods".

[٣٠] H. PADALKO, V. CHOMKO, S. YAKOVLEV, and D. CHUMACHENKO, "ENSEMBLE MACHINE LEARNING APPROACHES FOR FAKE NEWS CLASSIFICATION АНСАМБЛІЕВІ," Radioelectronic and Computer Systems, no. 4, pp. 5–19, 2023, doi: 10.32620/REKS.2023.4.01.

[٣١] N. Fahad et al., "Stand up Against Bad Intended News: An Approach to Detect Fake News using Machine Learning," Emerging Science Journal, vol. 7, no. 4, pp. 1247–1259, Aug. 2023, doi: 10.28991/ESJ-2023-07-04-015.

[٣٢] M. Alfreihat, O. S. Almousa, Y. Tashtoush, A. Alsobeh, K. Mansour, and H. Migdady, "Emo-SL Framework: Emoji Sentiment Lexicon Using Text-Based Features and Machine Learning for Sentiment Analysis," IEEE Access, vol. 12, pp. 81793–81812, 2024, doi: 10.1109/ACCESS.2024.3382836.

[٣٣] C. G. Kim, Y. J. Hwang, and C. Kamyod, "A Study of Profanity Effect in Sentiment Analysis on Natural Language Processing Using ANN," Journal of Web Engineering, vol. 21, no. 3, pp. 751–766, 2022, doi: 10.13052/jwe1540-9589.2139.

[٣٤] D. Singh, S. Shivprakash Barve, and A. Krishna Dwivedi, "OptiASAR: Optimized Aspect-Based Sentiment Analysis of Reviews With BiLSTM-GRU and NER-BERT in Healthcare Decision-Making," IEEE Access, vol. 13, pp. 47459–47474, 2025, doi: 10.1109/ACCESS.2025.3549303.

[٣٥] M. Kasri, M. Birjali, M. Nabil, A. Beni-Hssane, A. El-Ansari, and M. El Fissaoui, "Refining Word Embeddings with Sentiment Information for Sentiment Analysis," Journal of ICT Standardization, vol. 10, no. 3, pp. 353–382, Aug. 2022, doi: 10.13052/jicts2245-800X.1031.

[٣٦] W. Shahid, Y. Li, D. Staples, G. Amin, S. Hakak, and A. Ghorbani, "Are You a Cyborg, Bot or Human?-A Survey on Detecting Fake News Spreaders," IEEE Access, vol. 10, pp. 27069–27083, 2022, doi: 10.1109/ACCESS.2022.3157724.

[٣٧] H. Himdi, N. Zamzami, F. Najar, M. Alrehaili, and N. Bouguila, "Arabic Fake News Dataset Development: Humans and AI-Generated Contributions," IEEE Access, vol. 13, pp. 62234–62253, 2025, doi: 10.1109/ACCESS.2025.3556376.

[٣٨] A. M. K. Shehata, M. Nasser Al-Suqri, N. Eldin Mohamed Elshaiekh Osman, F. Hamad, Y. Nasser Alhusaini, and A. Mahfouz, "ArabFake: A Multitask Deep Learning Framework for Arabic Fake News Detection, Categorization, and Risk Prediction," IEEE Access, vol. 12, pp. 191345–191360, 2024, doi: 10.1109/ACCESS.2024.3518204.

[٣٩] P. Durga and D. Godavarthi, "Deep-Sentiment: An Effective Deep Sentiment Analysis Using a Decision-Based Recurrent Neural Network (D-RNN)," IEEE Access, vol. 11, pp. 108433–108447, 2023, doi: 10.1109/ACCESS.2023.3320738.

[٤٠] J. He, A. Wumaier, Z. Kadeer, W. Sun, X. Xin, and L. Zheng, "A Local and Global Context Focus Multilingual Learning Model for Aspect-Based Sentiment Analysis," IEEE Access, vol. 10, pp. 84135–84146, 2022, doi: 10.1109/ACCESS.2022.3197218.

[٤١] J. Alghamdi, Y. Lin, and S. Luo, "ABERT: Adapting BERT model for efficient detection of human and AI-generated fake news," International Journal of Information Management Data Insights, vol. 5, no. 2, Dec. 2025, doi: 10.1016/j.jjimei.2025.100353.

[٤٢] L. Sun and H. Wang, "Topic-Aware Fake News Detection Based on Heterogeneous Graph," IEEE Access, vol. 11, pp. 103743–103752, 2023, doi: 10.1109/ACCESS.2023.3318483.

[٤٣] J. Dougrez-Lewis, E. Kochkina, M. Liakata, and Yulan. He, "Knowledge Graphs for Real-World Rumour Verification," in Joint International Conference on Computational Linguistics, Language Resources and Evaluation, ELRA Language Resource Association, 2024.

[٤٤] J. Yu, J. Jiang, L. M. S. Khoo, H. L. Chieu, and R. Xia, "Coupled hierarchical transformer for stance-aware rumor verification in social media conversations," 2020. doi: 10.18653/v1/2020.emnlp-main.108.

[٤٥] S. Nkhata, Gibson; Mai, Quan; Oyshi, Uttamasha Anjally; Gauch, "Verifying Rumors via Stance-Aware Structural Modeling," 2025. doi: 2512.13559.

[٤٦] C. O. Truica, E. S. Apostol, R. C. Nicolescu, and P. Karras, "MCWDST: A Minimum-Cost Weighted Directed Spanning Tree Algorithm for Real-Time Fake News Mitigation in Social Media," IEEE Access, vol. 11, pp. 125861–125873, 2023, doi: 10.1109/ACCESS.2023.3331220.

[٤٧] A. H. J. Almarashy, M. R. Feizi-Derakhshi, and P. Salehpour, "Enhancing Fake News Detection by Multi-Feature Classification," IEEE Access, vol. 11, pp. 139601–139613, 2023, doi: 10.1109/ACCESS.2023.3339621.

[٤٨] D. Wang, W. Zhang, W. Wu, and X. Guo, "Soft-Label for Multi-Domain Fake News Detection," IEEE Access, vol. 11, pp. 98596–98606, 2023, doi: 10.1109/ACCESS.2023.3313602.

[٤٩] B. Liu, A. Wang, and C. Xia, "Interpretable Chinese Fake News Detection With Chain-of-Thought and In-Context Learning," IEEE Access, vol. 13, pp. 117186–117197, 2025, doi: 10.1109/ACCESS.2025.3571497.

[٥٠] T. Huang, Z. Xu, P. Yu, J. Yi, and X. Xu, "A Hybrid Transformer Model for Fake News Detection: Leveraging Bayesian Optimization and Bidirectional Recurrent Unit," Mar. 2025, [Online]. Available: http://arxiv.org/abs/2502.09097