IJICTR
International Journal of Information & Communication Technology Research

Volume 17- Number 3 – 2025 (34-45)

An In-Depth Comparative Analysis of Fake News Detection Approaches in Social Media: Methodologies, Challenges, and Future Directions

Tala Tafazzoli
ICT Research Institute (ITRC)
Tehran, Iran
tafazoli@itrc.ac.ir

Abouzar Arabsorkhi*
ICT Research Institute (ITRC)
Tehran, Iran
abouzar_arab@itrc.ac.ir

Received: 12 22 December 2024 – Revised: 15 January 2025- Accepted: 01 February 2025

Abstract—**The proliferation of fake news on social networks poses significant challenges for trust, security, and societal well-being. In this paper, we present a comprehensive study of fake news detection approaches and techniques, introducing a novel framework for news construction comprising four elements: news content, news context, news propagation, and news environment. We propose a new taxonomy of fake news detection techniques categorized into two primary types—individual methods (content-based, context-based, and propagation-based) and frameworks (hybrid and perception-aware methods). We highlight their strengths, weaknesses, and applicability by analyzing 14 state-of-the-art detection methods across platforms such as Twitter, Facebook, and Sina-Weibo. Furthermore, we address critical research gaps by identifying future directions, including early fake news detection, unsupervised learning, multimodal datasets, adversarial attacks on algorithms, multi-lingual platforms, and AI-generated content detection. Our findings and recommendations aim to serve as a foundation for developing new robust, scalable, and impactful fake news detection systems.**

Keywords: Fake news detection, Content-based methods, Context-based methods, Propagation-based methods, Hybrid methods, Perception-aware frameworks.

Article type: Research Article

Publisher: ICT Research Institute

* Corresponding Author

I. INTRODUCTION

Information and communication technology (ICT) has undergone significant changes from its beginning. There have been tremendous developments in ICT advancements that are changing user needs and requiring more efficiency and effectiveness [1]. ICT has undergone five stages of evolution. The first stage was the telegraph and telephone revolution in the 19th century. The rise of computing happened in the second stage of the revolution in the middle of the 20th century. The early computers were large and expensive; however, there were smaller, faster, and cheaper computers after development. In the third stage, the internet emerged and created the World Wide Web and social media in the late 20th century, where the rise of mobile computing happened in stage 4, leading to new industries and business models. Finally, stage 5 was the emergence of artificial intelligence.

Social media and mobile computing development have brought about different advantages and disadvantages. Major social networks include Facebook, WhatsApp, Twitter, and Sina Weibo. The main advantages of social networks are education, business, news dissemination, quick access to information and research, marketing tools, and social communications. One of the main disadvantages of social networks is the fast dissemination of misinformation and fake news. Social networks connect different nodes, so misinformation and fake news can spread to different nodes. Cybersecurity and trust are the other challenges of social networks. Misinformation can severely influence trust among users of a social network.

Enhancing social networks resulted in increased communication and news propagation on these platforms. The news which propagates on these platforms is either real or fake. There are different social network platforms such as WhatsApp, Facebook, Twitter, Sina Weibo, and the news social media. On online social networks, people share information, videos, and audio. Although these platforms are perfect for sharing information, fake news also disseminate on them. The spread of misinformation on online social networks causes users to believe the misinformation. Therefore, detecting fake news in online social networks (OSNs) is necessary. These platforms have central rumor-reporting centers.

Automatic detection of misinformation is a complex problem as it requires the model to understand how related or unrelated the information is compared to real information. Different reviews and surveys on fake news detection techniques compare them from different viewpoints. Hu et al. [2] prepared a survey on fake news detection algorithms from the perspective of characteristics of fake news, such as intentional creation, heteromorphic transmission, and controversial reception. Based on these characteristics, they have proposed three categories for fake news detection: intentional feature-based, propagation-based, and stance-based approaches. Phan et al. [3] reviewed the state and challenges of using GNNs for fake news detection systems and provided a GNN taxonomy. Their taxonomy categorizes fake news detection systems into content-based, multi-label learning-based, context-based, propagation-based, and hybrid-based systems. Shan et al. [4] classified fake news detection systems into four categories: content-based approaches, including knowledge-based, style-based, and multimodal-based approaches; propagation-based approaches, including news cascade, propagation graph approaches; and source-based approaches, including news author-based and social media user-based approaches.

As shown in Fig 1, we propose the construction of news in four categories: (i) news content, (ii) news context, (iii) news propagation, and (iv) news environment. News construction provides the building blocks of the news lifecycle. News Content focuses on the style of the content or knowledge about the content of the news. News Context refers to the context information of the news, which involves three basic entities, i.e., publishers, news pieces, and social media users, to determine whether it is credible or potentially misleading. News propagation is the news route on which the news spreads, and the news environment is divided into the micro-environment and the macro-environment. In the macro-environment, the news released at a time interval is considered, while in the micro-environment, the relevant news in that time interval is considered. The news environment is the external news environment in which the news is created and disseminated. In this environment, we consider the recent media opinion and attention to the news. The macro-environment contains the recent news items, and the news micro-environment is the subset of similar news items to the current news. The news content considers the internal relationships of the news content and linguistics. For example, consider a news item that propagates in social networks. Its content considers the linguistic features of the news, such as expressions, meaning, etc. Its context includes publishers, news pieces, social media users, such as the publisher's ID, and the users who have received the news piece. Its propagation includes its route for those who have received it, forwarded it, or like it, and all the routes on which the news has propagated.

We also propose a categorization for fake news detection techniques and describe each element in this category. As provided in Fig. 2, we categorize fake news detection techniques into individual methods and frameworks. Individual methods include content-based, context-based, and propagation-based methods, while frameworks include hybrid frameworks and News environment perception-aware fake news detection frameworks.

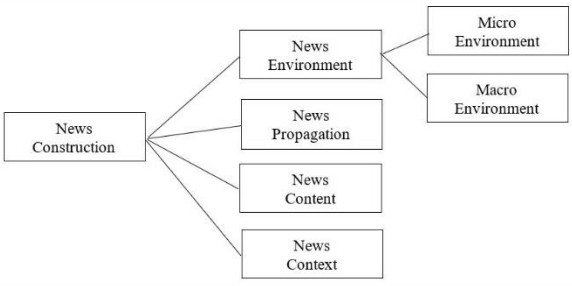

Figure 1. The categorization for news construction

Based on the studies performed on 30 papers, we categorize fake news detection techniques and propose a definition for each method. (i) Content-based fake news detection methods use the information in the news content and employ different techniques, such as knowledge graphs or machine learning techniques, to identify fake news. (ii) Context-based fake news detection techniques often utilize natural language processing and machine learning algorithms to assess the reliability of news content by considering not just the content itself but the broader context in which it appears. (iii) Propagation-based methods are based on news cascades or graphs built on social connections. (iv) Hybrid methods are content-context, content-propagation, and context-propagation, while some methods try to detect fake news using the perception of the news environment. (v) Perception-aware detection tries to extract environmental perception using similarity and deep learning methods. We study and discuss the strengths and weaknesses of the methods and show that each method applies to each social network.

This paper contributes to the field of fake news detection with the following key highlights:

- We propose a novel framework for news construction that captures the lifecycle of fake news in terms of content, context, propagation, and environment.
- A new taxonomy of fake news detection techniques is introduced, offering a unified categorization that bridges gaps in the existing literature.
- Through a meta-analysis of 14 state-of-the-art methods, we provide a comparative analysis of their applicability, strengths, and weaknesses across platforms and datasets.
- We identify significant research gaps and propose future directions, emphasizing early detection, unsupervised approaches, multimodal datasets, and AI-generated content challenges.
- We recommend strategies for building robust and scalable fake news detection systems that can adapt to real-world scenarios and adversarial attacks.

In Chapter II, we study the methodology of our research. Following, we discuss the definitions in Chapter III. Chapter IV presents a review of the relevant literature while Chapter V provides a comprehensive analysis of our approach and technique. At last, in Chapter VI, we discuss future directions, while Chapter VII concludes this research.

II. METHODOLOGY

This research analyzes the applicability of approaches and techniques for detecting fake news in social networks. Therefore, current research is applied research from the perspective of the goals. This research uses a qualitative and meta-analysis approach [5].

Meta-analysis is a method that combines and integrates the results of previous research in a specific field. This method works so that it shows the current state of knowledge (strengths, weaknesses, and challenges in using techniques) in a particular field, solves untrustworthiness, and identifies circumstances that need more research and studies. There are two meta-analysis methods: quantitative and qualitative.

We used the qualitative meta-analysis method to gain access to the current research goals. Based on Mish's opinion, a qualitative meta-analysis involves assembling, breaking, and examining the findings [6]. These activities detect properties and elements, construct a phenomenon, and convert the results to a new idea, changing the initial results to achieve new concepts. The current research considers social networks with different architectures and characteristics for content analysis. Besides that, we will analyze a wide range of approaches and techniques for identifying fake news with conditions of use, limitations of use, and various functions and capabilities. Finally, we will consider the strengths and weaknesses of the identified approaches and techniques. Considering their prerequisites, we identify the applicability or non-applicability of each fake news detection technique in social media.

Our literature review first reviewed fake news detection techniques available categories and taxonomies. Then, we searched the chosen fake news detection techniques, such as context-based, content-based, propagation-based, hybrid, and perception frameworks for fake news detection techniques. Between papers available on the internet, we chose highly ranked papers. From the 30 papers we reviewed at the beginning, we chose 14 papers for this study.

III. DEFINITIONS

A. An Introduction to Social Networks

Facebook relies on social graphs that represent the relationships between entities. The Facebook platform is the set of services, tools, and products that social networking services provide [7] [8]. There are billions of photos on Facebook, which are stored with four different resolutions, resulting in 4*N different photos on the Facebook platform. Therefore, performance is crucial in this system [9]. In this social network, individuals are shown as vertices, and relationships between individuals are shown as edges [10].

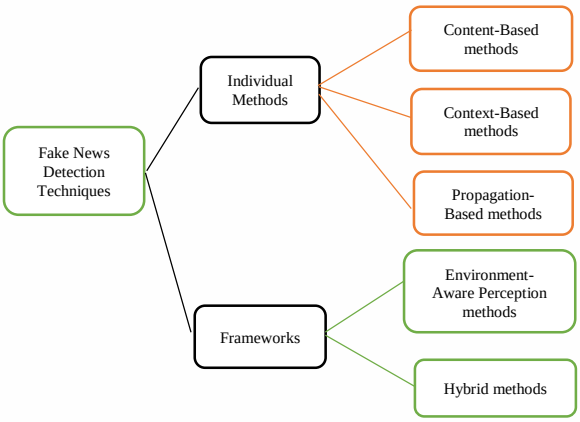

Figure 2. The categorization of fake news detection techniques

WhatsApp consists of multiple layers, including the client, business logic, and databases. The client layer handles user interactions, while the business logic layer processes messages and communications between users. The database layer stores user information, messages, and account metadata. Overall, the architecture is divided into the server and client components, adhering to the client-server architecture [11]. Furthermore, WhatsApp uses Signal, which provides an End-to-End Encryption (E2EE) protocol for communication.

Twitter [12] is a free social networking platform where users broadcast their posts, known as tweets. These tweets may contain text, videos, photos, or links. It uses graph databases and caching layers to allow users to follow each other. Launched in 2006, Twitter also enables users to curate their online experience, including what they see, which users and companies to follow, and what topics to search for. The user can do different tasks on Twitter, e.g., search, follow, post, and engage. One of the primary risks of Twitter is spreading fake news or misinformation, while Twitter bots, negative comments, data security, and privacy are other security risks.

Sina Weibo [13] is a Chinese microblogging website like Twitter and Facebook. It is widely used across China to share, disseminate, and receive information. Sina Weibo uses Alibaba Cloud, a public cloud platform, as a serverless computing platform to store and serve content, such as uploaded photos.

Fact-checking is defined as the process of verifying the accuracy of a statement or report. It could be done automatically or manually [14].

B.Fake News Definition

News is defined as a report of an event. Allcott and Gentzkow first defined fake news as false information used to mislead the audience [15]. Generally, information can be classified as follows: Fake news is not authentic and is used to cause harm. Disinformation refers to deliberately fabricated or false information that is propagated with deceptive intent to mislead. Misinformation is misleading information that spreads without the intent to cause harm or mislead. False information refers to inaccurate or incorrect information spread without malicious intent, despite lacking authenticity. Malinformation is genuine information that is disseminated with the intent to mislead.

C. Fake News Detection

Fake news detection aims to learn a function $f(.)$ that uses different types of information to determine whether a message m_i is fake or real.

$$f(.) = \begin{cases} 1 & if\ m_i\ is\ fake\ news \\ 0 & otherwise \end{cases} \quad (1)$$

Based on our methodology discussed in Section II, we chose 14 papers in 2 categories and five subcategories. These papers are discussed and analyzed in the following section. Specifically, we categorize fake news detection techniques into five subcategories:

(1) content-based fake news detection,

(2) context-based fake news detection,

(3) propagation-based fake news detection,

(4) hybrid fake news detection, and

(5) perception-aware fake news detection.

Content-based fake news detection assesses news intent using the features derived from news content. Context-based techniques, on the other hand, utilize the social context surrounding news objects on social media. Propagation-based techniques use news characteristics such as propagation path/tree, while hybrid methods combine any of the mentioned methods. Ultimately, perception-aware frameworks extract perception-aware features to detect fake news. We have discussed the core concepts and definitions used in the paper in Appendix I.

IV. FAKE NEWS DETECTION TECHNIQUES AND ALGORITHMS

A. Content-Based Fake News Detection

Pan et al. [16] propose innovative enhancements to TransE, B-TransE, and hybrid models for content-based fake news detection. They are the first to propose an approach utilizing positive and negative knowledge graph embedding, which is composed of entities and relationships between them. An article database refers to a collection of news articles containing a title, content, and an annotation. They use tools to generate knowledge graphs by utilizing a set of news articles and then use the transE or binary TransE models and hybrid approaches. More specifically, they utilize the knowledge graph to train a TransE model and compute a bias for classification purposes. In the B-TransE model, two models are trained on fake and true news. They perform their experiments on Kaggle's "Getting Real about Fake News" dataset alongside the BBC, Sky, and The Independent's news datasets. This study reports that the B-TransE model performs better than the TransE model. The precision of their method is 0.85. The hybrid approach performs well and has high precision, recall, and accuracy.

Verma et al. [17] suggest a content-based fake news detection approach known as WELFake, a two-phase benchmark model utilizing word embedding (WE) on linguistic features for fake news identification through machine learning classification techniques. Word Embedding is a representation of words in a lower-dimensional space. They have prepared a dataset with 72134 news items from Kaggle, McIntire, Reuters, and BuzzFeed. The data is preprocessed to omit missing, inconsistent, and irrelevant data. For feature engineering, linguistic features are extracted, including text-based linguistic features in two syntactic and semantic categories. At this step, essential features are chosen for data classification. To make features ready for machine learning algorithms, word embedding algorithms such as Term Frequency-Inverse Document Frequency (TF-IDF) and one-hot encoding, Word2Vec, GloVe, and FastText are employed. TF-IDF reflects the frequency of a word within a document through vectors, One-hot encoding converts categorical variables into binary matrix representations, and GloVe vectors are used for word representations. They detected fake news using SVM, NB, KNN, DT, Bagging, and AdaBoost. Finally, they employ ensemble learning to collect

outputs from various models and generate an output to minimize error and verification. The accuracy of their method is 96.73%.

Hu et al. [18] perform a deep analysis to determine the potential of LLMs. They use four prompting methods to ask the LLM to detect fake news. This study demonstrates that short language models, such as BERT, outperform even the best LLMs employed in the evaluation. Furthermore, they provide a comprehensive analysis of LLM-generated rationales and discovered that they are suitable for some uses, while concluding that LLMs are not considered a good substitute for fine-tuned short language models. Thus, they got the LLM rationale and input it as a news analysis to pre-trained Bert. They use the Chinese dataset Weibo21 and the English dataset GossipCop in addition to using the GPT3.5 turbo as LLM and BERT for short language models and Zero-Shot, Zero-Shot CoT, Few-Shot, and FewShot-CoT for prompting.

B. Context-based fake news detection

Yan et al. [19] propose a context-based fake news detection method that utilizes GNN and Graph Attention Network. They consider the news and its relevance within the broader social context. In their study, the researchers constructed a diverse graph consisting of news articles, comments, and users. Using meta-paths, they deconstructed this complex graph into two subgraphs - one focusing on news-comments and the other one on news-users. Following this, they utilized GCN to extract features from these subgraphs and implemented a Graph Attention Network to learn feature representations of nodes within each subgraph. Ultimately, they introduce an attention mechanism between the two subgraphs to combine the node representations for fake news classification. This study uses the FakeNewsNet dataset for its evaluations and proposes a model with an accuracy of 90.16.

Gue et al. [20] suggest an innovative fake news detection model for mixed languages by incorporating a multiscale transformer to capture the semantic information present in the text. Due to the distinct principles governing languages and their vocabularies, there is a need for more methods for effectively processing multi-language texts. The authors of this study have introduced a fake news detection model – MST-FaDe – based on a multiscale transformer designed for mixed languages. They take the Chinese-English mixed scenario and use embedding strategies for the two types of characters. Thereafter, representative vectors from both languages are merged into a standardized representation format. In this context, the breadth and horizontal relationships of the transformer model are extended to enhance its performance in mixed-language scenarios. They have transformed a fake news detection problem into a binary classification problem. This study uses the Sina Weibo dataset for its evaluations, supplemented with English texts and news articles, and reports an accuracy of 0.88 for their proposed model.

Sitaual et al. [21] propose a credibility-based detection of fake news. They aim to recognize common signs that indicate the credibility of news by examining both the source and the content to distinguish fake news. Their research shows that distinguishing fake news from real news can be achieved by analyzing aspects of the source and the content. They also note that although certain features differ between fake and true news, they do not necessarily enhance the ability to predict fake news accurately. This study uses a total of 26 features, including factors such as the number of authors, credibility of coauthors, historical records, and sentiments within content. They obtained 26 features. Various classification algorithms were tested, with Adaboost emerging as the most effective classifier for these features. The average F1 score obtained by source credibility features is 0.77, while the average F1 score for content credibility is 0.68.

C. Propagation-based fake news detection techniques

Julio et al. [22] suggest an automated approach for creating a scoring model to assist fact-checking organizations in identifying fake news within images disseminated on WhatsApp. Their tool integrates with a fact-checking tool such as WhatsApp Monitor. They have suggested a novel ranking system that considers the prevalence of fake news and is compatible with the mentioned tools. They employed various features to detect fake news in images circulated on WhatsApp. Furthermore, they gathered features from various aspects, including content (such as textual and image properties), source (like the publisher's identity), and environmental factors. They extracted 181 features for fake news detection. They used SVM, MLP, and XGBoost [4] for ranking, and achieved a 95% confidence interval.

Hu et al. [23] propose CompareNet, a model designed to compare news to the external knowledge base to detect fake news. In CompareNet, the authors build a directed heterogeneous document graph containing topics and entities, utilizing the connections between sentences, topics, and entities. They enhance the external knowledge base, and the entities connect the knowledge base and the news document. They studied the content represented by a knowledge-based entity with an entity comparison network. Finally, they identify fake news by associating the features with the representation of the news document. They used a directed heterogeneous document graph and extended a heterogeneous graph attention network to learn the representations of news and entities. Furthermore, CompareNet employs LSTM to encode a sentence and get its feature vector, using the softmax function with the attention vector and attention weights to normalize across the neighboring nodes. They have also calculated the type-level attention weights based on the current node embedding and the type embedding using the softmax function. After L-layer graph convolution, they finally get all the node representations aggregating neighborhood semantics. They extract structural embedding using TransE and textual embedding to have knowledge-based entity representations. Finally, they integrate structural and textual embedding using gating integration. To assess the effectiveness of their approach, the researchers compared news documents with knowledge bases and calculated a comparison vector, utilizing two news datasets: (i) SLN and (ii) LUN. They report a micro F1 score of 89.17% for SLN and 69.05% for the LUN dataset, which demonstrates that CompareNet improves the results compared to other state-of-the-art methods.

Monti et al. [24] introduce a novel automated method for detecting fake news, using geometric deep learning – a generalization of non-Euclidean deep learning that has proved effective in analyzing heterogeneous data – that operates on graph-structured data. The proposed model operates in a supervised fashion and relies on a substantial amount of labelled data. The researchers incorporate four types of features into their model: (i) user profile, (ii) user activity, (iii) network and spreading, and (iv) content features. GCN has superseded traditional CNNs on grids, as they perform permutation-invariant aggregation on the neighborhood of a vertex within a graph. Spectral graph CNNs operate by harnessing Laplacian eigenvectors and the traditional Fourier Transform. Various permutation-invariant aggregation operators exist, with the Laplacian operator being one of them. They employed a four-layer GCN architecture with two convolutional layers. The model produces a 64-dimensional feature map as output and includes two fully connected layers that generate 32 and 2-dimensional output features, respectively. They utilized one graph attention head in each convolutional layer within their model. The Scaled Exponential Linear Unit (SELU) was consistently employed as the non-linearity function across the entire network. In the input generation phase, they paired each URL with tweets that mentioned it, and the URL is subsequently represented as a graph in their model. News propagation occurs from one node to another if one node follows the other in the graph representation used by the model. The model defines spreading trees for news diffusion paths by considering two key parameters: (i) the timestamp of retweets and (ii) the social connections between the nodes in the graph. Both nodes and edges in the graph have features associated with them to describe their characteristics within the model. The features associated with edges represent one of the relations: (i) following, (ii) news spreading, or (iii) both directions within the model. This study reports a high level of accuracy, with an ROC AUC score of 92.7% in detecting fake news using their proposed model.

Barnabò et al. [25] present two main contributions in their paper. First, they examine active learning (AL) strategies within the context of GNNs, particularly for the purpose of detecting misinformation. They later introduce Deep Error Sampling (DES), a novel deep active learning framework incorporating uncertainty sampling, leading to superior performance compared to traditional AL strategies. Active learning is a machine learning technique that allows the model to interactively query the user to obtain the desired outputs rather than relying solely on a predetermined dataset for training. It initiates with a small dataset and generates predictions for the remaining data. The model can identify a subset of data for which it is uncertain about its predictions and request the user to provide the correct labels for further training. The labelled data obtained can subsequently be updated to enhance the model's performance. Additionally, they have introduced an innovative deep learning approach called DES, which offers enhanced performance when combined with uncertainty sampling. Active learning assists in efficiently and effectively identifying unknown items for labelling, which can often be a costly process. This approach helps maximize performance. Various active learning strategies include (i) random sampling, (ii) uncertainty sampling, and (iii) diversity sampling. Deep active learning strategies are also designed to work effectively with deep learning models. They have also introduced the deep error sampling method that constructs the embedding of samples from the network input with the intent to determine if the classifier misclassifies new samples. They have conducted experiments with GNN-based misinformation detection approaches, such as GCN, GAT, and GraphSAGE, that operate on news diffusion graphs. The trained model accepts a graph as input, representing the diffusion cascades of each URL on a social network like Twitter or Facebook. Their method is supervised. They use two datasets, FbMultiLingMisinfo and PolitiFact, that include diffusion cascades, and report 89% accuracy of the proposed method for GraphSAGE with 100 iterations on the FbMultiLingMisinfo dataset. It achieves almost 92% AUC for 20 iterations on the PolitiFact dataset.

D. Hybrid-based fake news detection

Raza et. Al. [26] propose a model based on transformer architecture, composed of two primary components: (i) an encoder that learns representations from fake news data and (ii) a decoder that predicts future behavior based on past observations. This research introduces a new framework based on transformer architecture to learn representations from fake news by utilizing information from news content and social contexts to classify the data. The proposed model preserves a temporal order in the sequences, where each word in the sequence is temporally arranged and assigned a timestamp. This process implies that the first few words correspond to different timesteps, such as 0 and 1. The news ecosystem comprises three fundamental entities. They also introduce a classification model called FND-NS (Fake News Detection through News content and social context) leveraging bidirectional and autoregressive transformers (BART) for a novel task. However, the researchers have made modifications to BART, as it accepts one piece of information as input. In contrast, their proposed model takes a rich set of features from news content and social context into the encoder. They have also used multi-head attention to weigh the importance of multi-head pieces of information. This approach allows the model to assign greater weight to posts with higher interactions, emphasizing more influential posts. Additionally, they have modified Bart's data loader. More specifically, the second difference is how the next token is predicted. They utilize token-level tasks and, in the final step, apply a linear transformation together with a SoftMax layer for the classification task. The model is evaluated using the NELA-GT-19 [23] dataset, which consists of news articles from various sources, and Fakeddit [24], a multimodal dataset containing texts and images extracted from Reddit posts. They use weak supervision for label creation, allowing labels to be created at the source level and used as proxies for the articles. The accuracy of their method is 0.748, surpassing the performance of other methods.

Lu et al. [27] developed a pioneering Graph-aware Co-Attention Network (GCAN) model to forecast the source tweet’s authenticity and detect potentially

suspicious retweeters. Their approach incorporates brief text content along with the sequence of retweets from users and user profiles as input. The model identifies fake news under three distinct scenarios:

1) based solely on the short text content of the source tweet,

2) excluding user comments, and

3) disregarding the network structure of the social network and the diffusion network. The proposed GCAN consists of five key components: (i) user characteristics extraction, (ii) news story encoding, (iii) user propagation representation, (iv) dual co-attention mechanisms, and (v) prediction-making. They used Twitter15 and Twitter16 as their datasets, while reporting an accuracy of 0.87 on the Twitter15 and 0.9 on the Twitter16 datasets.

E. Environment-aware perception fake news detection frameworks

Sheng et al. [28] present a comprehensive framework that considers both the environment and the content of the news. The news environment is divided into (i) the micro-environment and (ii) the macro-environment. The macro-environment provides a global perspective on what the mass audience reads and focuses on, while the micro-environment focuses on the distribution of news items related to similar events. The authors observe two critical signals from the environment:

(1) popularity and

(2) novelty.

Popularity is defined as the degree of similarity between a news item and a previously established popular item. In the micro news environment, the news items focus on the topic, while the news gives new information about that topic. The perceived vector of the environment measures the similarity between the news item and the environment without much information loss. To assess this similarity, the authors calculate the cosine similarity between the news and other items in macro-environments. In contrast the assessment consists of Gaussian Kernel Pooling followed by concatenation and output normalization in micro-environments. Novel news content is considered an outlier in the surrounding environment. This study uses a multi-layer perceptron in micro-environments. The authors combine the results of environment-perceived vectors with the fake news content detectors for prediction purposes, where fake news detection was achieved via (i) content-based fake news detection (such as BERT or EANN) or (ii) knowledge-based fake news detection methods (namely DeClae and MAC). Finally, they use a classifier (MLP) for prediction. This study uses the Chinese Weibo dataset along with English news databases for the detection of fake news and reports a maximum achieved accuracy of 0.831.

Fang et al. [29] present a fake news detection framework to judge the authenticity of news content and post context in both micro and macro environments. Their approach consists of three primary components: (i) news semantic environment construction, (ii) news semantic environment perception, and (iii) prediction. They divide the news environment into both macro and micro environments. The internal relationship between the posts and the semantic context is explored in macro environment. The semantic environment is formed by collecting news items shared within a defined time window T prior to the release of the target news item, and cosine similarity is in turn utilized to extract similar news items. The authors note that fake news has novelty, while true news is real data and tends to be consistent with prior accurate information. A BERT model is used to capture the semantic features through its unique word vectors, followed by the utilization of a GCN to extract semantic patterns and implicit evidence to identify misinformation in the macro environment. On the other hand, the process differs in the micro-environments. The micro semantic environment is constructed by selecting the top r news items most relevant to the target post, in order to identify contradictions in the news. Using an attention mechanism, the semantic correlations between inputs are modeled. Finally, the authors concatenate (i) the macro semantic environment perception feature, (ii) the micro semantic environment feature, and (iii) the output of a fake news detector to determine the authenticity of the target post. This study uses a Chinese dataset of mainstream media sources along with an English dataset of news headlines, collected from various news media sources. They combined their method with baseline content-based and knowledge source-based fake news detection methods and reported improved accuracy in English and Chinese datasets. They achieved early fake news detection while using their framework with content-based or knowledge source-based fake news detectors.

V. THE ANALYSIS OF OUR APPROACH AND TECHNIQUE

In this study, we have investigated 30 papers and selected 14 for further analysis. We assessed these papers based on their methodology, datasets, and the accuracy of their results. Additionally, we explored the applicability of these methods in the context of social networks. Finally, we discussed the strengths and weaknesses of each approach.

In Table 1, we present a comprehensive comparative analysis of different fake news detection methods based on their categories, underlying algorithms, and the platforms to which they are applicable. We have examined 14 methods encompassing five distinct categories:

(1) Content-based fake news detection,

(2) Context-based,

(3) Propagation-based,

(4) Hybrid, and

(5) Environment-based methods.

Among the methods studied, five methods effectively detect fake news on Twitter; two on Facebook, and one on WhatsApp. Additionally, one method works successfully on the multilingual SinaWeibo dataset, while another operates separately on Chinese Weibo21 and English GossipCop datasets. Furthermore, nine of the methods are applied to news-based datasets. Our study delves into each method's strengths and

challenges. To conduct our analysis, we used the meta-analysis method. We aimed to select the latest and highly ranked methods within each category based on their citation counts and publication years.

The methodologies include supervised classification-based machine learning algorithms, such as logistic regression and support vector machines, that are used by [21] [22] [17] across Twitter, Facebook, WhatsApp, and news datasets, as well as in fact-checking contexts. Graph-based methods, such as GCN and GAT, are utilized in studies [19][27][23][24][25] on Twitter, Facebook, and news datasets along with fact-checkers. Transformer-based methods are featured in studies [26] [20] on news datasets and Sina Weibo. Lastly, study [18] employs LLM and SLM on the Weibo21 dataset and GossipCop, while studies [29] [28] utilize perception-aware frameworks. Our review indicates that graph-based methods are more widely used than other approaches and demonstrate strong accuracy and F1 scores.

In Table 1, we discuss each method's strengths and challenges. Ten of the 14 analyzed methods are supervised, while two are unsupervised and do not need labeled data [25] [26]. Five methods can detect fake news at early stages [24] [26], and three methods [20] can handle bilingual datasets.

When discussing the challenges associated with these methods, it is important to note that all the supervised methods are unable to detect new and real data. Additionally, utilizing transformers for bilingual detection can lead to increased cost and complexity [20]. There are also privacy concerns while working with WhatsApp data [22].

In [16], the authors have generated a knowledge graph, analyzed it with B-TransE, and incorporated bias functions. This approach achieved an accuracy of 0.9, demonstrating strong performance even under conditions of incomplete data.

For each method, we demonstrate its precision, recall, and F1 score. However, some of the studied methods do not report these metrics, making direct comparisons difficult since the experiments were performed on different datasets and under varying circumstances. A future benchmark study can facilitate a fair comparison by evaluating these methods under the same circumstances using the same datasets.

VI. FUTURE DIRECTIONS

Excellent progress has been made in fake news detection. However, there are still flaws that should be addressed in future work, which will be discussed here.

Robust Detection: Fake news detection techniques should accurately identify fake news and resist adversarial attacks on the detectors and other kinds of attacks. For example, attackers may inject adversarial content into fake news detectors. Therefore, fake news detection methods should be developed to detect fake news robustly.

Real-world Scenarios / Changes in datasets: Fake news detection techniques should be able to detect and work correctly when the dataset changes or when using data from real-world scenarios and big datasets.

Early detection: Fake news detection techniques should be able to detect fake news in situations where it is not spread widely and at an early stage.

Unsupervised/Semi-Supervised approaches: Since gathering labelled datasets is difficult and methods that work in a supervised manner cannot detect new scenarios, it is necessary to develop unsupervised and semi-supervised fake news detection methods with good performance and accuracy.

LLMs and AI-generated content: Until recently, most content and news were generated by humans, but now a lot of it is generated by AI and LLMs [30]. While LLM has progressed, the detection of AI-generated content is more complex. Some research has been conducted on machine-generated text, but most concentrates on detecting machine-generated text rather than the factuality of its content. In the future, AI and LLMs will be used for news generation. The next generation of fake news detection techniques should be able to detect machine/human-generated fake/real news and adapt to the era of LLMs. On the other side, LLMs should be used for fake news detection. Research [18] shows that human and Short Language Models outperform LLM-based fake news detectors.

Multimodal or real datasets: Current datasets used in fake news detection research are text, and there is a need for real fake news datasets and datasets that include audio, video, and text simultaneously.

Limited research has been conducted on **adversarial attacks** targeting techniques for the detection of fake news.

VII. CONCLUSIONS

In this study, we have enhanced the understanding of fake news detection by presenting a comprehensive framework for news construction and a novel taxonomy for categorizing detection techniques. Our comparative analysis of 14 state-of-the-art methods highlights their strengths, weaknesses, and applicability across social networks and news platforms. By addressing critical research gaps – such as early detection, unsupervised methods, and the detection of AI-generated content – we lay the groundwork for future innovations in this field.

Furthermore, we advocate the development of multimodal and real-world datasets to enhance the robustness and scalability of detection methods. Our findings emphasize the necessity of building systems resilient to adversarial attacks and adaptable to ever-evolving challenges of misinformation detection. As social networks and AI-driven content generation continue to grow, these contributions will be crucial in combating fake news and safeguarding trust in digital communication.

TABLE I. AN ANALYSIS OF THE APPROACHES IN FAKE NEWS DETECTION

Approach	Ref.	Technique	Ref.	Use Case	Strengths	Challenges
Content Based	[2] [3] [4]	TransE, B-TransE, hybrid	[16]	News [16]	Able to process incomplete imprecise knowledge graphs [16], Supervised method/ labeled data[4], TransE simplicity and efficiency [33]	Difficulty in validating extracted triples [16], Difficulty of detecting fake news with the absence of pre-existing knowledge graph for a given topic [16], Knowledge graph generation and scalability complexity, TransE weakness in complex relationships
		GPT3.5 turbo, BERT	[18]	Weibo21, GossipCop	Comprehensive empirical analysis of LLMs in fake news detection, Multi-perspective rationale generation by LLMs, novel hybrid detection framework (ARG and ARG-D), Superior performance over baselines	Complexity and diversity of fake news, Limitations of small language models, Underperformance of Large Language Models in direct detection, Integration of LLMs and SLMs, Cost and efficiency
		SVM, NB, KNN, DT, Bagging and AdaBoost	[17]	News	Novel two-phase model, Ensemble learning approach	Feature selection for fake news detection, Inability to detect new unseen patterns/ real world scenarios [2],
Context Based	[2] [3] [4]	GCN, GAT	[19]	Twitter, News	Novel heterogeneous graph framework, Meta-Path-based subgraph decomposition, Integration of GCNs and attention mechanisms [19]	Dependence on social context data availability, Scalability and computational complexity, Sensitivity to noisy or malicious user data, Cold start and early detection limitations, Generalizability across different platforms [19]
		Multiscale transformer	[20]	Sina Weibo	Addresses mixed-language scenarios, Innovative multiscale transformer architecture, Empirical validation on real-world data	Lack of fine-grained error analysis, Computational complexity, Limited discussion on early detection, Potential overfitting to dataset
		Assess credibility	[21]	Buzzfeed, Politifact	Novel use of source credibility, Empirical analysis of author information, Combination of content and source features, Statistical validation	Limited dataset size and diversity, Dependence on author metadata, Limited early detection capability, No deep learning or advanced modeling utilized
Propagation Based	[2] [3] [4]	Graph Attention Network	[23]	News	Integration of external knowledge, Heterogeneous graph architecture, Topic-enriched representations, Empirical performance	Dependency on knowledge base quality, Computational complexity
		SVM, MLP, XGBoost	[22]	WhatsApp, Fact checker	Practical application, Efficiency gains, Comprehensive feature engineering, Dataset innovation	Context-specific limitations, Feature complexity, Interpretability gaps, Scalability concerns
		GCN	[24]	Twitter	High accuracy, Early detection capability	Limited metric reporting, Platform specificity, Data collection complexity, Potential for dataset bias
		Active Learning, GNN	[25]	Facebook, Twitter	Novelty in active learning for misinformation, Reduction in human labeling effort, Robustness and reliability	Limited reporting of precision, Recall, and F1 score, Emphasis on Twitter-like data, Complexity of GNNs, Active learning overhead
Hybrid	[2] [3] [4]	Transformer	[26]	News	Early fake news detection, Integration of content and social contexts, Transformer-based architecture, Weak supervision for labeling, Comprehensive evaluation	Dependence on social context data, Complexity and computational cost, Potential label noise
		GCAN	[27]	Twitter	Handles short text & sparse data, Explainability via dual co-attention, Integration of multi-modal features, Robust performance	Reliance on user metadata, Computational complexity, Limited generalizability, Qualitative explainability

Environment perception frameworks	[28] [29]	GCN, BERT	[29]	Chinese and English news	Innovative semantic environment modeling, Early detection capability, Advanced deep learning techniques	Computational complexity, Potential for overfitting
		Cosine similarity, MLP, Gaussian Kernel Pooling	[28]	Weibo, English news	Novel "Zoom-Out" approach, Dual environment analysis, Compatibility with existing models, Weak supervision & real-world applicability	Dependence on mainstream news quality, Computational overhead, Limited exploration of base detectors

TABLE II. PRECISION, RECALL, ACCURACY OF EACH METHOD

	Algorithm	Precision	Recall	F1 Score
[16]	B-TransE	0.85	0.80	0.83
[18]	ARG-D	-	-	0.790
[17]	Welfake	0.967	0.968	0.967
[19]	GCN, GAT	0.921	0.914	0.917
[20]	Multiscale transformer	0.930	0.927	0.928
[21]	Credibility	-	-	0.8
[22]	SVM, MLP, XGBoost	-	-	-
[23]	Graph Attention Network	-	-	75.2% (LIAR), 82.3% (FakeNewsNet)
[24]	GCN	-	-	-
[25]	GNN	-	-	-
[26]	Transformer	0.93	0.92	0.92
[27]	GCAN	0.85	0.83	0.83
[28]	Cosine similarity, MLP, Gaussian Kernel Pooling	0.83	0.84	0.84
[29]	-	0.874 (Chinese)	0.861 (Chinese)	0.867 (Chinese)

VIII. APPENDIX I - CORE CONCEPTS IN THIS PAPER

In this appendix, we discuss about machine learning methods used in this paper.

Traditional machine learning algorithms:

Support Vector Machine (SVM): SVM is used for binary classification. The basic idea is to find a hyperplane that separates the d-dimensional data into two classes.

K-Nearest Neighbors (KNN): The KNN classifier assigns each observation to the most similar labeled example and calculates the Euclidean distance between them. Another issue is the decision about the number of neighbors for a node.

Decision Tree (DT): A decision tree has multiple key components. The root node is at the top, making the initial decision point. Internal nodes represent a choice based on the task. The leaf nodes are used for predictions or decisions.

Adaboost: It assigns weights to the ML algorithm's original training set and then adjusts the weights after each learning phase.

Bagging: It generates classifiers if the base algorithm is unstable in case of significant changes in the classifier caused by minor changes to the training input.

Deep learning methods:

Multi-Layer Perceptron (MLP): This is a fully connected network. It is a feed-forward artificial neural network. It has an input layer, an output layer for decision-making, and one or more hidden layers. It makes decisions based on the output layer.

Convolutional Neural Networks (CNN): CNN is a discriminative deep learning architecture that learns from input. It does not need human feature extraction. It has multiple convolutions and a pooling layer on it, each with a certain number of parameters.

Recurrent Neural Network (RNN): This uses sequential or time series data and feeds the previous layer's output as input to the current stage. RNN learns from training input and combines input and output with information from previous input.

Long-Short-Term Memory (LSTM): This is a popular form of RNN, with some units having a vanishing gradient problem. A memory cell in LSTM stores data for an extended period of time, and three

gates manage the flow of information into and out of the cell.

Graph-based methods:

TransE model: Given a training set S of triplets (h, l, t) in which we have two entities $h, t \in E$ (the set of entities) and a relationship l ∈ L (the set of relationships), the TransE model learns the vector embedding of the entities and relations. To learn the embedding, it minimizes a margin-based ranking criterion over the training set:

$$L = \sum_{(h,l,t) \in S} \sum_{(h',l,t') \in (S'_{(h,l,t)})} [\gamma + d(h + l, t) - d(h' + l, t')]_+ \quad (2)$$

, where S' is the set of corrupted triples.

Graph Neural Networks (GNNs): GNNs are a class of deep learning methods designed to infer the data described by graphs. They can be directly applied to graphs and provide an easy way to perform node-level, edge-level, and graph-level prediction tasks.

Graph Attention Networks (GATs): GATs are a variant of Graph Neural Networks (GNNs) that leverage attention mechanisms for feature learning on graphs. GATs assign an attention coefficient to each neighbour, indicating the importance of that neighbour's features for the feature update of the node.

GraphSAGE: This inductive framework uses node feature information to generate node embeddings for unseen data. It is a powerful GNN capable of scalable learning on graph-structured data and makes inferences for unseen nodes by aggregating unsampled local frameworks.

Generative models:

LLM: A large language model is an artificial intelligence algorithm that uses deep learning techniques and enormous datasets to understand, summarize, generate, and predict new content. The term generative AI is also closely connected with LLMs, which are, in fact, a type of generative AI specifically architected to help generate text-based content. LLMs learn tasks using prompts that contain instructions.

- Zero-Shot Prompting: This prompt contains a task description and given news.
- Zero-Shot CoT Prompting: This is a simple chain-of-thought approach.
- Few-shot Prompting provides news labels and task-specific prompts.
- Few-Shot CoT prompting demonstrates the reasoning step and provides data labels.

Transformer: A transformer model is a neural network that learns context and, thus, meaning by tracking relationships in sequential data like the words in this sentence. Transformers leverage self-attention mechanisms to weigh the importance of different words in a sentence, allowing for parallel processing and capturing long-range dependencies in data.

REFERENCES

[1] Darshana Chathuranga Mihidukula. Evolution of ICT Applications, 2023. URLhttps://darshanamihidukula.medium.com/evolution-of-ict-applications-44893e281fdf. Accessed on 2023-04-05.

[2] Bo Hu, Zhendong Mao, and Yongdong Zhang. An overview of fake news detection: From a newperspective. Fundamental research.

[3] H. T. Phan, N. T. Nguyen, and D. Hwang. Fakenews detection: A survey of graph neural networkmethods.Applied Soft Computing, 139.

[4] Y. Shen, Q. Liu, N. Guo, J. Yuan, and Y. Yang.Fake news detection on social networks: A survey.Applied sciences, 13.

[5] T. A. Trikalinos, Salanti G., Zintzaras E., andJ.P.A. Ioannidis. Meta-analysis methods.Advances in genetics, 60.

[6] G. C. Mish.Websterss ninth new collegiate dic-tionary. Springfield: Merriam-Webster, 1989.

[7] wikipedia. Facebook Platform, https://en.wikipedia.org/wiki/Facebook_Platform, Accessed on 2024-6-12.

[8] wikipedia. Social graph, https://en.wikipedia.org/wiki/Social_graph, Accessed on 2024-6-12.

[9] H. Barrigas, D. Barrigas, M. Barata, P. Furtado, and J. Bernardino. Electronic sortition. In Proceedings of the International Conference on In-formation Systems and Design of Communica-tion, pages 173–176, 2014.

[10] J. Ugander, B. Karrer, L. Backstrom, and Mar-low C. The anatomy of the facebook social graph.

[11] K. Lewi, C. Rain, S. Weis, Y. Lee, H. Xiong,and B. Yang. Scaling backend authentication atfacebook.IACR Cryptology ePrint Archive.

[12] Jerry Wallis. WhatsApp Tech Stack Ex-plored — The Tech Behind Series, 2024. URL https://intuji.com/whatsapp-tech-stack-explored/. Accessed on 2024-12-06.

[13] Amanda Hetler. Twitter, 2023. URLhttps://www.techtarget.com/whatis/definition/Twitter#:~:text=Twitter%20is%20a%20free%20social,or%20website%2C%20Twitter.com. Accessed on 2023-12-06.

[14] A. Vlachos and S. Riedel. Fact checking: Taskdefinition and dataset construction. InProceed-ings of the ACL Workshop on Language Tech-nologies and Computational Social Science, pages 18–22, 2024.

[15] Wikipedia. Weibo, 2023. URL https://en.wikipedia.org/wiki/Weibo. Accessed on 2023-12-06.

[16] J.Z Pan, S. Pavlova, C. Li, N. Li, Y. Li, andJ. Liu. Content based fake news detection using knowledge graphs.The Semantic Web, 11136:669–683, 2018.

[17] P. K. P. K. Verma, P. Agrawal, I. Amorim, and Prodan R. WelFake: Word embedding over linguistic features for fake news detection. IEEE Transactions on Computational Social Systems, 9:881–893, 2021

[18] B. Hu, Q. Sheng, J. Cao1, Yang Li Y. S.,D. Wang, and P. Qi. Bad actor, good advisor:Exploring the role of large language models infake news detection.Association for the Ad-vancement of Artificial Intelligence,.

[19] F. Yan, M. Zhang, B. Wei, W. Jiang, and K. Ren.Fake news detection utilizing social context information with graph convolutional networks and attention mechanisms. In EITCE'23: Proceedings of the 2023 7th International Conferenceon Electronic Information Technology and Com-puter Engineering, pages 406 – 413, 2024.

[20] Z. Guo, Q. Zhang, F. Ding, X. Zhu, K Yu, A Novel Fake News Detection Model for Context of Mixed Languages Through Multiscale Transformer, IEEE Transactions on Computational Social Systems, 2023.

[21] N. Sitaula, C. K. Mohan, J. Grygiel, X. Zhou,and R. Zafarani. Credibility-based fake news detection, disinformation, misinformation, and fake news in social media, pp 163–182, 2020.

[22] Julio C. S., P. M. Reis, F. Belém, F. Murai,J. M. Almeida, and F. Benevenuto. Helping fact-checkers identify fake news stories shared through images on whatsapp. In Web Media '23:

Proceedings of the 29th Brazilian Symposium on Multimedia and the Web, page 159–167, 2023.

[23] L. Hu, T. Yang, L. Zhang, W. Zhong, D. Tang,C. Shi, N. Duan, and M. Zhou. Compare to the knowledge: Graph neural fake news detection with external knowledge. InProceedings ofthe 59th Annual Meeting of the Association forComputational Linguistics, page 754–763, 2021.

[24] F. Monti, F.and Frasca, D. Eynard, D. Mannion,and M.M. Bronstein. Fake news detection on social media using geometric deep learning, arXiv:1902.06673, 2019.

[25] G. Barnabò, F. Siciliano, C. Castillo, S. Leonardi, P. Nakov, G.D.S. Martino, and F. Silvestri. Deep active learning for misinformation detection using geometric deep learning. Online Social Network Media, 33:11, 2023.

[26] S. Raza and C. Ding. Fake news detection based on news content and social contexts: a transformer-based approach. International Journal of Data Science and Analytics.

[27] Y. J. Lu and C. T. Li. Gcan: Graph-aware co-attention networks for explainable fake news detection on social media, arXiv:2004.11648, 2020.

[28] Q. Sheng, X. Cao, J.and Zhang, R. Li, D. Wang, and Y. Zhu. Zoom out and observe: News environment perception for fake news detection. Association for Computational Linguistics. 2022.

[29] X. Fang, H. Wu, Y. Jing, J.and Meng, B. Yu, H. Yu, and H. Zhang. Nsep: Early fake news detection via news semantic environment perception. Information processing and management, 61:17, 2024.

Tala Tafazzoli received her Ph.D. degree in Computer Engineering from Amirkabir University of Technology, Tehran, Iran. She is an Assistant Professor at ICT Research Institute (ITRC). Her research interests include blockchain technology, deep learning, and cybersecurity. She has more than 25 years of research experience in E-commerce, and Cyber Security.

Abouzar Arabsorkhi received his Ph.D. in Information Systems Management from the University of Tehran. He is a faculty member at ICT Research Institute. With over 20 years of research experience, he focuses on Security Management, Risk Management, Security Architecture, and Prototype Certification. His main research interests include the Internet of Things, Blockchain, and Emerging Technology Security.