A Review of Fake News Detection Techniques for Arabic Language

Taghreed Alotaibi[1], Hmood Al-Dossari[2]
Computer Science Department, Imam Mohammad Ibn Saud Islamic University, Riyadh, Saudi Arabia[1]
Information Systems Department, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia[1, 2]

***Abstract*—The growing proliferation of social networks provides users worldwide access to vast amounts of information. However, although social media users have benefitted significantly from the rise of various platforms in terms of interacting with others, e.g., expressing their opinions, finding products and services, and checking reviews, it has also raised critical problems, such as the spread of fake news. Spreading fake news not only affects individual citizens but also governments and countries. This situation necessitates the immediate integration of artificial intelligence methodologies to address and alleviate this issue effectively. Researchers in the field have leveraged different techniques to mitigate this problem. However, research in the Arabic language for fake news detection is still in its early stages compared with other languages, such as English. This review paper intends to provide a clear view of Arabic research in the field. In addition, the paper aims to provide other researchers working on solving Arabic fake news detection problems with a better understanding of the common features used in extraction, machine learning, and deep learning algorithms. Moreover, a list of publicly available datasets is provided to give an idea of their characteristics and facilitate researcher access. Furthermore, some of limitations and challenges related to Arabic fake news and rumor detection are discussed to encourage other researchers.**

Keywords—Fake news detection; rumors; classification; Arabic language

I. INTRODUCTION

The influence of traditional information sources, such as television and newspapers, and how users gather and consume news has diminished in comparison to earlier times. The expansion of social media platforms has been a key factor in this transition. Social media platforms are another kind of technological innovation. Users can access platforms such as Twitter, Facebook, and Instagram to build their profiles, share opinions, interact with others having shared interests, and facilitate cycles of cooperation. Over the years, more users have shown an interest in using social media. According to Kepios, there were 4.76 billion social media users around the world in 2023, which equates to 59.4% of the world's total population [1]. The reason behind this spike in the use of social media is that social media platforms are designed to be more attractive and highly suitable for social communication. Social media has also become a way for companies and governments to reach the people by providing news, showcasing their services, giving updates, or launching marketing campaigns. However, some limitations arise from using social media, such as the spread of fake news and rumors, as well as spamming. Many individuals who use social media platforms to stay in touch with friends and family also use these platforms to find news and information. According to a report from the Pew Research Center, 48% of adults in the US use social media as a news source [2].

Fake news is a major problem that started early, attracting significant attention in 2016 during the US presidential elections. Different fake news items and rumors usually appear during special events and cover different domains, such as those related to elections, natural disasters, or health, such as the coronavirus disease 2019 (COVID-19) pandemic. A study on Twitter showed that false news spreads more quickly and reaches 100 times as many readers as true news [3]. Therefore, many researchers are working to solve this problem, ranging from analyzing the types of fake news [4] to trying to find the most effective method for detection [5].

In the Arab regions, different rumors and false information have been spread during the COVID-19 pandemic [6]. Rumors and false information have also proliferated in politics, as was the case during the Arab Spring and the Syrian crisis [7]. The Arabic language is considered one of the most commonly spoken languages in the world, and an essential language for Muslims worldwide, who numbered about 1.8 billion in 2015 and are expected to increase to three billion in 2060, according to the Pew Research Center [8]. The Arabic language presents various challenges. First, there are different dialects used in different Arab countries, as well as from regions in the same country. The language also has a rich vocabulary that leads to the occurrence of misleading information from different dialects, making it more difficult for the system to detect [9]. In social media, the complexity of understanding the Arabic language is increased because users on social media, such as Twitter, use two forms of the language: Modern Standard Arabic (MSA), which is the formal language, and Dialect Arabic (DA), which is informal, used in daily communication between users, and is more common than MSA [10].

This paper working to focus in two aspects: features extraction techniques and the datasets used. This is because while classification algorithms play a critical role in fake news detection, their effectiveness heavily relies on the quality and relevance of the features provided to them [11]. Feature extraction techniques enable the models to capture the subtle nuances and contextual cues that differentiate fake news from real news [12]. By incorporating these features, classification algorithms can leverage the rich information present in the data, leading to improved accuracy and robustness in fake news detection. Majority of studies in Arabic fake news detection have predominantly focused on the application of specific features extractions techniques to identify the veracity of news with no concern on their limitation[13][14][15]. However, this study surpasses mere technical application by shedding light on the strengths and weaknesses of each technique.

On the other hand, focusing on identifying and analyzing available datasets is crucial for advancing research in fake news detection. It allows researchers to improve data quality, benchmark models, and collaborate effectively to address the growing challenge of misinformation in our digital world. There is a need for research which reviews the available dataset in Arabic for fake news detection. For that, this research coming to fill this gap by analyzing available public Arabic datasets according to their domain, size, labels, annotation method, source and the features used. We believe this work provide researchers in the field with valuable insights into the available dimensions to initiate their research endeavors. The contributions of this review paper are as follows:

1) Investigate research in Arabic fake news detection by selecting studies that cover different features for detecting fake news and utilize publicly available datasets.

2) Provide a clear insight into the available helpful techniques used for feature extraction in detecting fake news, in general, and in the Arabic language, in particular;

3) Investigate the publicly available Arabic datasets and show their properties, including a list of the available Arabic fact-checking websites that help build datasets;

4) Explore the limitations and prospective avenues concerning datasets, utilized features, and classification methods associated with detecting Arabic fake news.

The rest of this paper is arranged as follows. Section II describes a brief background of fake news, providing its definitions, and impacts. Section III identifies the different approaches for feature extraction. Section IV investigates the works conducted in the field of Arabic fake news detection. Subsequently, Section V provides a list of the available Arabic fact-checking websites used. Section VI investigates the available Arabic dataset in fake news detection. Finally, Section VII discusses some of the limitations and future direction, followed by the conclusion in Section VIII.

II. Background

A. Fake News Definition

"Fake news" does not have a consistent definition. Scholars define the term differently based on the purpose of their research. For example, one study [16] defined fake news as "a news article that is intentionally and verifiably false." In 2018, a European Commission report defined "fake news" as "all forms of false, inaccurate, or misleading information designed, presented, and promoted to cause public harm intentionally or for profit" [17]. Research in [18] has provided and defined a set of terms related to fake news, including hoax, rumor, spam, misinformation, disinformation, clickbait, satire, propaganda, and hyperpartisan. *Hoaxes* are facts that are either erroneous or inaccurate and are presented as actual facts in news stories [19]. They consist of fabricated information that has been purposefully created to appear true, and because they involve highly intricate and large-scale fabrications, they frequently cause substantial material harm to the victim [20]. *Rumor* is information initiated by a potentially untrustworthy person and circulated among users before it can be verified to be true, false, or unconfirmed [21]. *Spam* refers to any irrelevant or unsolicited messages sent by individuals or groups on social media, such as advertisements, malicious links, or content of poor quality [22]. *Misinformation* is false information that is spread unintentionally, for example, by mistake or updating specific knowledge without intentionally misleading [23]. *Disinformation* is a type of false information spread among users with the main goal of misleading others for some purpose, such as deception of some person [19] or promoting a biased agenda [24]. The concepts of misinformation and disinformation often confuse readers. While both refer to inaccurate or fake information, disinformation is created with malicious intent, whereas this is not necessarily the case with misinformation [25][19]. *Clickbait* is a story whose title or news headline is different from the content itself; it is mainly employed to attract users to access a specific website to increase traffic and, consequently, boost revenue [26][24]. *Satire* is a noteworthy literary tool employed in creating news articles, serving the purpose of both criticism and amusement for readers [27]. This form of discourse often incorporates a substantial amount of irony [28]. *Propaganda* encompasses a form of persuasive communication that employs unidirectional messaging to influence the attitudes, emotions, perspectives, and behaviors of specific target audiences, driven by ideological, political, and religious motives [29][30]. *Hyperpartisan* refers to heavily biased or one-sided narratives [31], particularly in the political arena, denoting strong partiality toward individuals, parties, circumstances, or events [32]. This review will use fake news and false news as a big umbrella for these concepts and will not consider each concept distinctly because there is an overlap in the techniques performed.

B. Impact of Spreading Fake News

Fake news influences users, leading them to accept biased or false beliefs, changing how they react to true news stories [12]. Spreading fake news not only affects individuals but also harms society. In 2016, fake news drew global attention during the US election [33], when a large amount of fake news was shared on Twitter. The spread of fake news goes beyond the political sector and also applies to other sectors. During the COVID-19 pandemic, a significant amount of fake news and rumors spread throughout the health sector. Approximately 80% of consumers in the US reported that they had read fake news during the pandemic [34]. In addition, during the fire disasters in Australia, some fake maps and pictures were shared on social media [35]. While this may have increased awareness of the disaster, this fake information may also have cost people their lives [35].

In the Middle East, Arabic countries have also been influenced by some fake news, for example, during the COVID-19 pandemic. According to a study on COVID-19 misinformation in Jordan, different Arabic media outlets promoted conspiracy theories regarding the pandemic, with the most prevalent ideas focusing on the virus's origins [36]. Most of the fake news spreading in Arab countries are related to the political sector, such as news related to the Arab Spring in 2010, Islamic State terrorist group (ISIS) campaigns, and the Beirut explosion of 2020 [37].

III. Features Used in Fake News Detection

Feature extraction and selection are techniques used in text mining that have shown effectiveness in enhancing performance in different tasks [10]. Having large features requires more powerful computation and enough memory; hence, there is a vital need to choose appropriate features. In fake news detection, researchers have recently applied feature extraction and selection with good results. In detecting fake news, three approaches are commonly used by researchers: knowledge-based (fact-checking), content approaches, and context approaches [38] [12]. This section will describe each feature in more detail for better understanding. Fig. 1 summarizes these features.

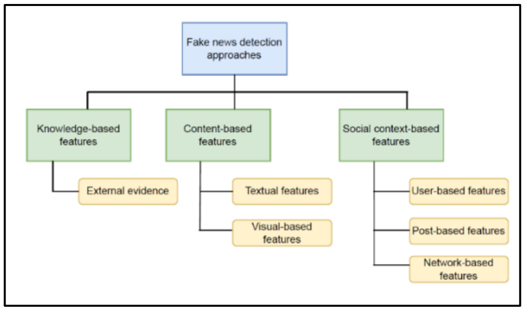

Fig. 1. Fake news detection approaches.

A. Knowledge-based Approaches

This approach works by attempting to discover whether there are facts that support the claim made in a news item [39]. In this approach, fact-checking strategies can be included, which entail finding documents and web pages that support a news item based on information retrieval methods [40]. Few works have used the web to search and retrieve evidence on search query formalization [40]. Study [40] performed a study combining social media conversations with evidence from external sources. The research developed dataset which contain social media conversations and external evidence related to each rumor. Their results showed that combining evidence with rumor is more effective than using rumors or evidence alone. Also, in [41] a system worked by incorporating external evidence from the web with some signals from language style.

B. Content-based Approaches

Content-based features focus on information that can be directly extracted from the text, such as linguistic features. The content features can be classified into (textual) linguistic-based [74] and visual-based [12].

1) Linguistic-based features: Linguistic-based features extract content from the text at different levels, such as characters, words, sentences, and documents [12]. In fake news detection, the common linguistic features used are lexical, syntactic, and semantic [42][12]. Lexical features, considered as the actual wording (usage) in the text, can be at a word level, such as the total number of words, word length, and the frequency of unique words, or at a character level, such as characters per word [12][43]. One of the commonly used features is bag-of-words (BOW), n-gram, which is used to find the most prominent word contents or expressions [44]. On the other hand, syntactic features include sentence-level features, such as the order of words in a sentence, grammar, and syntactic structure of a sentence [12]. The other features related to linguistics are semantic features, which include using the Natural Language Processing (NLP) technique to extract information, such as opinion mining, and sentiment analysis to extract emotions and opinions from texts [38]. Some research also extract topics in texts or posts on social media using Latent Dirichlet Allocation (LDA) [43]. Word embedding, an example of the distribution semantic technique, is useful in detecting fake news with machine learning and deep learning [45]. The linguistic feature is an effective indicator in detecting fake news during the initial propagation phase with the user features [46]. This method is simple but limited because it requires deep knowledge of the domain; thus, it is difficult to generalize [47]. Another limitation when dealing with linguistic features is that the features extracted from social media, such as Twitter, may not be enough for machine learning approaches because the text is short [48]. Moreover, it cannot be used for detecting fake news that only has images or videos [48].

2) Visual features: Visual content features extract features from visual elements, such as videos or photos, using deep learning techniques [49]. Fake news creators use individual vulnerabilities and images or any kind of visual cues to provoke an emotional response from users [12]. To extract visual features from images that can be used in detecting fake news, applications such as clustering scores and similarity distribution histograms are used [49] [12]. In addition, statistical features can be used in fake news verification, such as the multi-image ratio, number of images in an event, image ratio, and long image ratio [49] [12]. Other research has used visual features, such as the polarity of the image and the probability of the image being manipulated [50]. This kind of feature can be integrated with other features, such as text-based features and user-based features, which provide a good result [51]. Few researchers in fake news detection have investigated visual features due to the lack of availability of datasets containing images and video [12] [52]. In addition, the techniques for maintaining these features are more complex than those needed for other features[12] [50].

C. Social Context-based Features

Social context-based features are related to the information surrounding news, such as the user's characteristics, the reactions of other users to posts or news, and social network propagation features [42][12].

1) User-based Features: User-based features focus on the news publisher to evaluate the credibility of the source [12].

The features are used to confirm credibility and reliability for an individual user using demographics, registration age, number of tweets written by the user, number of followers or following, number of tweets authorized by the user, user photo, user sentiment, tweet repetition, and so on [43]. Most social media platforms provide users with a verification status after they provide information to the social media company to confirm their identity. This verification status is used by researchers in detecting a user's reliability [53]. When using user-based features, it is important to understand that the availability of this information is a critical concern due to privacy and access constraints on some platforms such as Twitter.

2) Post-based features: Post-based features can be used to identify the veracity of news from different aspects relevant to social media posts[12]. This includes analyses of user feedback, reactions, opinions, and responses, in general, as indicators of fake news. Some of these features include comments, likes, tagging, user ratings, sentiment, and emotional reactions [42]. Some research dealing with post-based features have dealt with a number of retweets, likes, shares, and others [54][55]. Another unique feature represents the social responses to a post by users who interact with the news story. These features rely on the wisdom of the crowd in detecting fake news and show its effectiveness in detection. In this situation, a few studies have analyzed responses from different perspectives, such as sentiment, emotion, and stance toward news items [12]. In fact, [56] used users' response information as core input data in detecting fake news. They believed that users' responses have rich information that researchers can benefit from.

3) Network-based features: Social media users construct networks grounded in relationships, interests, and subjects [12]. Detecting fake news necessitates extracting network-based features to unveil and represent discernible patterns suitable for identification. Different types of networks can be constructed. Shu et al. [12] categorized them into stance network, where the nodes represent tweets related to tweets and the edges represent weights of stance similarities [57] [58]; co-occurrence network, which counts users' written posts related to the same news article [47]; and friendship network, which is based on the following and followers of the users who posted in relation to tweets [59]. After building networks, some matrices can be used for feature representation, such as degree and clustering coefficients, to represent the diffusion network[60]. Research using network-based features is limited compared with other features because of the complexity that can emerge when analyzing the patterns [46]. In addition, finding a dataset that contains enough information for a network is difficult patterns [46].

Research in the field uses each feature alone or incorporates both content and contextual features in detecting fake news [47]. These methods are promising regarding effectiveness but are challenging when relying only on one type of feature in automatic fake news detection. Table I summarizes the shortcomings and merits of each feature.

TABLE I. SHORTCOMINGS AND MERITS FOR FAKE NEWS DETECTION APPROACHES

Approaches		Strength and weakness
Knowledge-based approach	***External evidence***	(+) Usually Improve the performance[40]. (+) Enhanced interpretability[98]. (-) Need information retrieval techniques[40]. (-) Need strategies for checking the credibility of sources.
Content-based approaches	***Textual-based***	(+) Simple method. (+) Appropriate for early detection of fake news[46]. (-) Requires deep knowledge of the domain [47]. (-) May not fully convey the message's context well as underlying intent, leading to misunderstandings [18].
	Visual-based	(+) Enhanced the performance[51]. (+) May provide contextual information[50]. (-) Limited dataset containing visual content, especially in Arabic language[12] [52]. (-) Need complex techniques to be maintained from professional & more time for analysis[12][50].
Social context-based features	***User-based***	(+) Appropriate for early detection of fake news[46]. (-) There are some privacy concerns.
	Post-based	(+) Appropriate for early detection of fake news[46]. (-) Need more investigation[12].
	Network-based	(+) Good for identifying influential users [12]. (-) Provide low performance in the early stage of detection[46] [48]. (-) Complex to analyzed and require enough information about structure and users' connection [46]. (-) Need advanced computational techniques and expertise in network analysis[12].

IV. FAKE NWES DETECTION AND ARABIC LANGUAGE

Fake news detection in Arabic has gained significant attention recently but is still in its infancy. Some studies were conducted using a simple technique, such as [61], which suggested a solution based on using the rule-based model. The dataset was adapted from Almujaiwel[1], where they built a dictionary that contains a set of fake news with keywords for each news. The rule-based model worked by checking the primary and secondary keys in the dictionary, whether they are in a tweet text or not. In the context of traditional machine learning, various researchers have used common supervised machine learning, such as [62]. The study used word frequency, count vector, and Term Frequency - Inverse Document Frequency (TF-IDF) as features. Moreover, two word-embedding methods were used: FastText and Word2Vec. To collect tweets, the research used Infection Disease Ontology. The result after testing on Logistic Regression (LR), Naïve Bayes (NB), and Support vector machine (SVM) classifiers showed the best accuracy, with 84% for the LR model with a count vector. Alkhair et al. [63] performed another study to detect the content of fake news on YouTube. The study collected data related to rumors about three famous people in Arab countries (topic of the dataset), namely, Fifi Abdu, Adel Imam, and Abdelaziz Boutaflika. The study used features related to content: n-grams with TF-IDF.

[1] https://github.com/salmujaiwel

For the classification process, SVM, Multinomial Naïve Bayes (MNB), and Decision Tree (DT) were used. The best accuracy and precision were registered for SVM across these topics. While the textual features showed their effectiveness, other studies have suggested using user features. El Balloul et al. [64] proposed a model called (CAT) for the credibility analysis of Arabic content on Twitter. CAT is built using a combination of features related to content and user. It has 26 content-based features and 22 user-based features. The research constructed a dataset from Twitter in Arabic, which is considered topic-independent. CAT was trained using NB, SVM, and RF. CAT registered a higher weighted average F-measure of 75.8% using RF. Overall, sentiment was found to be a highly crucial feature in defining credibility, especially the negative sentiment, as well as the URL in the author profile linked to their website. Mouty and Gazdar [65] conducted their research using the same datasets and features of [64]. To enhance the accuracy of the classifier, the study developed an algorithm for discovering the similarity between username and display name in a Twitter account and the similarity score between tweets and Google search results. For classification, they used RF, DT, SVM, and NB. The combination of user/content features and the new two-similarity score features registered 78.71% accuracy for the RF model.

Using similar features, Jardaneh et al. [55] conducted another experiment using LR, Adaptive Boosting (AdaBoost), Random Forest (RF), and DT for classification. The dataset used was from [7]. The research confirmed that the sentiment feature has a beneficial effect on the system’s accuracy, especially when ensemble-based machine learning algorithms are employed. Taher et al. [66] performed a study that employed a Harris Hawks Optimizer (HHO) for the feature selection approach. Briefly, the researcher used a combination of features related to user profile, content-based, and linguistic features, namely, TF-IDF, BOW n-grams, and Binary Term Frequency (BTF). Eight machine learning algorithms were tested: eXtreme Gradient Boosting (XGB), NB, K-Nearest Neighbor (KNN), linear discriminant analysis (LDA), DT, LR, SVM, and RF. LR was selected with HHO algorithms, which performed well as the wrapper feature selection approach registered a 5% increase compared with [55], where the accuracy was 82%.

Alzanin and Aqil [67] performed another study for detecting rumors using unsupervised and semi-supervised expectation maximization (EM). The dataset used contained 271,000 tweets belonging to 88 non-rumor events and 89 rumor events, and 16 features related to users and content were used. For the classification tasks, the researchers compared their proposed model with the supervised Gaussian Naïve Bayes (GNB) model. The results showed the proposed model outperformed GNB with an accuracy of 78.6%. In the health sector, [68] focused on cancer treatment information disseminated via social media. The research extracted tweets annotated manually by experts in the field. The total number of datasets was 208 tweets. Given the small set of data, the researchers performed over-sampling to enhance the performance of the model. The features used were TF-IDF and n-grams, while the classifiers were SVM, LR, KNN, Bernoulli Naïve Bayes (BNB), Stochastic Gradient Decent (SGD), and j48. The research also used an ensemble method using RF, AdaBoost, and Bagging. The results confirmed that the over-sampling process enhanced the performance of all models of machine learning. Meanwhile, RF outperformed the others using 4- and 5-grams based on accuracy. In study [69], they also relied on a set of extracted features from user and textual features. To extract the textual features, they used both classic word embedding (word2vec, fastText, and Keras embedding layer) and context-based embedding (Multilingual Arabic Bidirectional Encoder Representations from Transformers (MARBERT) and ARBERT) with deep learning models[70]. Two deep learning schemes were used: Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (BiLSTM). The result showed that MARBERT with CNN provided the best accuracy of 95.6%. Alawadh et al. [71] performed another experiment using machine learning algorithms and Mini-BERT. The research used the dataset available from [72]. First, the research preprocessed the dataset by applying standard text 2 numeric encoding. Subsequently, DT, NB, RF, linear support vector (LSV), and mini-BERT were applied with three separate splits using the holdout validation technique (70/30, 80/20, 90/10). The result showed that mini-BERT exhibited consistent performance among the splits with increasing training data, while the machine learning classifiers showed varying performances across the splits. The highest accuracy registered with mini-BERT was 98.4%.

Numerous fake news was spreading during the COVID-19 pandemic. These encouraged researchers to build datasets covering this domain, and the deep learning approaches provided them with excellent performance. The researchers started using Neural Network (NN) and a set of language models, such as [9]. They performed experiments to detect misinformation spreading via Twitter related to the COVID-19 pandemic. The dataset size was 8,786 collected Arabic tweets. For classification, the study used eight of the traditional machine learning models, which were MNB, SVM, XGB, SGD, and RF. On the other hand, deep learning models were used, which are CNN, Recurrent Neural Network (RNN) and, Convolutional RNN (CRNN). Both experiments with machine learning and deep learning used different representation features, including word frequency and word embedding. The XGB showed the highest accuracy in detecting misinformation in this study. Study in [73] collected datasets about the COVID-19 pandemic from Twitter on the types of fake news and misinformation to detect fake news. The tweets were annotated in two ways: manual and automatic. For feature extraction, they used count vector, word-level TF-IDF, character-level TF-IDF, and n-gram-level TF-IDF. They chose the best classifier based on performance, which was trained in the manually annotated dataset to automatically annotate the remaining unlabeled dataset as false and real. The study used six classifiers, including RF, NB, LR, Multilayer Perceptron (MLP), XGB, and SVM. As a result, LR exhibited the best performance for both manual and automatic annotated datasets, which registered an F1-score of 93.3%. In this study, the stemming and rooting techniques failed to increase the performance of the classifiers.

TABLE II. SUMMARY OF ARABIC FAKE NEWS DETECTION STUDIES AND CORRESPONDING FEATURES AND MODELS USED.' E' EVIDENCE, 'T' TEXTUAL, 'V' VISUAL, 'U' USER, 'P' POST, 'N' NETWORK

Study	**Dataset**	**Feature type**						**Models**	**Platform**	**Feature Extraction Methods**	**Evaluation results**
		E	***T***	***V***	***U***	***P***	***N***				
Alotaibi and Alhammad[61]	2000 tweets.	-	✓	-	-	-	-	Rule-based	Twitter	Building dictionary.	Accuracy = 78.1 % Precision= 70% Recall =98%
Alsudias & Rayson [62]	2000 tweets	-	✓	-	-	-	-	LR, SVM, NB	Twitter	Word frequency, Count vector and TF-IDF, FastText, Word2Vec.	Accuracy= 84.03% Precision = 81.04% Recall = 80.03% F1-score= 80.5%
Alkhair et al[63]	3434 comments	-	✓	-	-	-	-	SVM, DT, MNB	YouTube	N-grams, TF-IDF	Accuracy= 95.35% Precison = 92.77% Recall= 83.12%
El Ballouli et al[64]	9000 tweets.	-	✓	-	✓	✓	-	NB, SVM, RF	Twitter	-	Precision = 76.1% Recall = 76.3%, F1-score = 75.8%
Mouty and Gazdar [65]	Dataset[64]	✓	✓	-	✓	✓	-	DT, SVM, NB, RF.	Twitter	-	Accuracy= 78.71% Precision = 78.5% Recall= 78.7%, F1-score = 78.5%
Jardaneh et al. [55]	Dataset[7]	-	✓	-	✓	✓	-	RF, DT, LR, AdaBoost	Twitter	-	Accuracy= 76% Precision= 79% Recall= 82% F1-score= 80%
Taher et al. [66]	Dataset[7]	-	✓	-	✓	✓	-	XGB, NB, KNN, LDA, DT, LR, SVM, RF.	Twitter	TF, TF-IDF, BTF, N-grams	Accuracy= 82% Precision = 82% Recall= 86%, F1-score= 84%
Alzanin and Aqil[67]	177 events.	-	✓	-	✓	✓	-	EM	Twitter	-	Accuracy=78.6% precision=79.8% recall=80.2% F1-score=78.6%
Saeed et al [68]	208 tweets	-	✓	-	-	-	-	SVM, LR, KNN, BNB, SGD, j48, Bagging, AdaBoost, RF	Twitter	TF-IDF, N-gram	Accuracy = 83.50% Precision= 86% Recall= 83% F1-score= 83%
Alyoubi et al.[69]	5000 tweets.	-	✓	-	✓	-	-	CNN, BiLSTM	Twitter	Keras Embedding Layer, word2vec, FastText, ARBERT. MARBERT, (word and sentence-level)	Accuracy = 95.6% Precision = 95.6% Recall = 95.6% F1-score = 95.6%
Alawadh et al.[71]	Dataset[72]	-	✓	-	-	-	-	DT, NB, LSV, RF, Mini-BERT	Articles	BERT	Accuracy = 98.43% Precision = 100% Recall =97.5% F1-score = 98.73%
Alqurashi et.al[9]	8786 tweets	-	✓	-	-	-	-	MNB, SVM, XGB, SGD, RF. CNN, CRNN, RNN	Twitter	TF-IDF (world-level, N-grams), FastText, word2vec,	Accuracy= 86.2% Precision =67% Recall= 25% F1-score = 37%
Mahlous and Al-Laith [73]	37029 tweets	-	✓	-	-	-	-	RF, NB, LR, MLP, XGB, SVM.	Twitter	Count vector, Word-level TF-IDF, N-gram-level TF-IDF. Char-level TF-IDF	**Manual dataset:** Precision=87.8% Recall=87.7% F1= 87.8% **Automatic dataset:** Precision= 93.4% Recall= 93.3% F1-score= 93.3%
Elhadad et al.[74]	COVID-19-FAKES [74]	-	✓	-	-	-	-	KNN, DT, LR, MNB, LSVM, BNB, Perceptron, NN, XGB, ERF, BME, GB, AdaBoost	Twitter	TF, TF-IDF, N-gram, char-level, word embedding	-
Haouari et al [75]	ArCOV19-Rumors [75]	-	✓	-	✓	✓	✓	MARBERT, AraBERT, Bi-GCN, RNN+CNN	Twitter	-	Accuracy= 75.7% macro-F1=74%
Ameur and Aliane [76]	AraCovid19-MFH [76]	-	✓	-	-	-	-	AraBERT, mBERT, Distilbert Multilingual, arabert	Twitter	AraBERT, mBERT, Distilbert Multilingual, arabert Cov19 , mbert	F-score= 95.78%

								Cov19 , mbert Cov19		Cov19	
Nassif et.al[77]	Kaggle [78] ,10000 records	-	✓	-	-	-	-	AraBERT, QaribBert-base, Araelectra, MARBERT. Arabic-Bert, ARBERT, RobertBase, GigaBert-base.	Articles & tweets	-	Accuracy= 98.5% Precison=99.1% Recall=98.2% F1-score=98.6%
Touahri and Mazroui [79]	200 claims, 3380 evidences	✓	✓	-	-	-	-	Scoring function	Tweets & articles	-	Accuracy= 92.7%. Precision= 54.66%. Recall = 56.13% F1-score= 55.2%
Elaziz et al.[80].	ArCOV19-rumors [75] OSACT4[81] Dataset [73] Dataset [82].	-	✓	-	-	-	-	AraBERT	Twitter & articles	MTL and AraBERT.	Accuracy=95.5% Precision = 96.2% Recall = 96.29% F1-score = 96.28%
Amoudi et al[83]	ArCOV19-Rumors [75]	-	✓	-	-	-	-	SVM, NB, KNN, DT, RF, SGD, LR, XGB, GRU, RNN, LSTM, Bi-RNN, B-GRU, Bi-LSTM	Twitter	TF-IDF, N-gram, AraVec	Accuracy= 80% Precision= 80% Recall= 72 F1-score= 75%
Al-Yahya et al. [45]	ArCOV19-Rumors [75] AraNews[84] ANS [85] COVID-19-Fakes [74]	-	✓	-	-	-	-	CNN, GRU, RNN, AraBERT, ArElectra, QARiB, ARBERT, MARBERT	Articles & Twitter	Word level, char-level, Word2Vec, Glove, FastText, doc2vec	Accuracy = 97.5% Precision = 95.6% Recall = 95.6% F1= 95.3%
Fouad et .al[86]	Dataset[87] 1980 tweets	-	✓	-	-	-	-	LSVC, SVC, MNB, BNB, SGD , DT, RF CNN, LSTM, CNN+LSTM, BiLSTM, CNN + BiLSTM	Articles & Twitter	Word embedding, N-gram	Accuracy= 83.92%
Shishah[88]	Covid-19-Fakes [74] ANS [85] Satirical[89] AraNews[84]	-	✓	-	-	-	-	BERT with joint learning	Articles & Twitter	-	Accuracy = 85% Precision = 86% Recall = 86% F1-score = 85%
Bsoul et al.[90]	2652 news records	-	✓	-	-	-	-	LR, RF, NB, SGD, NN, DT.	Twitter	TF-IDF, BOW.	Precision=84% Recall=78% F1-score= 81%
Saadany et al.[89]	Satirical Fake News[89]	-	✓	-	-	-	-	MNB, XGB, CNN	News articles	Count vector, TF-IDF, N-grams, char-level, word-level, FastText	Accuracy =98.59% Precision = 98.49% Recall = 98.61% F1-score = 98.49%
Khouja[85]	ANS[85]	-	✓	-	-	-	-	BERT, LSTM	News title	Word-level, charr-level	Precision = 64.1% Recall = 64.6% F1-score = 64.3%
Nagudi et al.[84]	AraNews [84] ATB [2] ANS[85]	-	✓	-	-	-	-	AraBERT, mBERT, XLM-R Base, XLM-R Larg	News articles	AraBERT, mBERT, XLM-RBase,, XLM-RLarg	Accuracy = 74.12% F1-score = 70.06%
Himdi et al[91]	1098 records	-	✓	-	-	-	-	SVM, NB, RF	Article	POS, Emotion, Polarity, linguistics (syntactic, semantic)	Precision = 79% Recall= 79% F1-score= 79 %
Albalawi et al.[52]	4025 tweets.	-	✓	✓	-	-	-	AraBERT (different version) ARBERT, MARBER, MARBERTv2, QARIB, Arabic Bert, Arabert Covid-19, mbert Covid-19 Ara-DialectBERT, AraT5, VGG-19, ResNet50	Twitter	AraBERT (different version), ARBERT, MARBER, MARBERTv2, QARIB, Arabic Bert, Arabert Covid-19 mbert Covid-19, Ara-DialectBERT, AraT5,VGG-19, ResNet50	Accuracy = 89.8% Precision = 89.47% Recall = 89.87% F1-score= 89.64%

[2] https://www.ldc.upenn.edu/collaborations/past-projects

Study in [74] collected and annotated a set of data from Twitter in Arabic/English language. The research was conducted in two phases. First, a binary classification model was trained on a set of collected ground-truth data, which were obtained from the official websites and the official Twitter accounts of the United Nations (UN), United Nations International Children's Emergency Fund (UNICEF), and World Health Organization (WHO). The second phase carried out the annotation for the, unlabeled tweets. The system used extraction features, such as TF, TF-IDF (n-gram, character level), and word embedding with 13 machine learning algorithms, including KNN, DT, LR, MNB, Linear Support Vector Machines (LSVM), BNB, Perceptron, NN, XGB, Ensemble Random Forest (ERF), Bagging Meta-Estimator (BME), Gradient Boosting (GB), and AdaBoost. Meanwhile, Haouari et al. [75] built the ArCOV19-Rumors dataset, which covered some claims about COVID-19. The research presented a benchmark on claim-level verification and tweet-level verification to exploit the content, user profiles, propagation structure, and temporal features. The Bidirectional Graph CN (Bi-GCN) and RNN+CNN, as well as the AraBERT and MARBERT models were tested on the dataset. The final results showed the best accuracy for AraBERT and MARBERT, with 73% and 75.7%, respectively.

Furthermore, [76] used a pre-trained transformer AraBERT, Multilingual BERT (mBERT), and Distilbert Multilingual under baseline transformer models. The study fine-tuned mBERT and AraBERT on [92]'s datasets containing dialects. The output from this process were two models called AraBERT COV19 and mBERT COV19, which were used in addition to the above-mentioned three models for the experiment (five models overall). The experiment found that the pretrained COVID-19 models were helpful after fine-tuning in detecting false information. Nassif et al. [77] used eight BERT transformer-base models. The research conducted two experiments: using the dataset from Kaggle[3], which was written in English and translated to Arabic using Google translator, and using another dataset, sourced from Twitter and newspaper agency websites. GigaBert-base and QARiB Bert-base provided the best result on the translated dataset and the collected dataset, respectively. Study in [79]focused on the task of claim verification, which involves a collection of claims and corresponding evidence in the form of text snippets sourced from web pages [93]. The researchers achieved an F1-score of 55.2% by employing a scoring function that evaluates the negation and concordance between the claims and their associated text snippets. The determination of negation and concordance levels was performed through a manual list-based approach. Some research focused on feature extraction and selection, such as [80], which used three main methods, including multi-task learning, a transformer-based model, and Fire Hawk Optimization (FHO) algorithm. Three datasets related to COVID-19 and the detection of fake news were used, which include ArCOV19-Rumors, as well as datasets from [73] and [82]. AraBERT was used for feature extraction via multi-tasking learning and fine-tuning approaches, a novel metaheuristic algorithm to select the most pertinent features from the contextual feature representations. The average accuracy in binary classification using these datasets was 91%. Amoudi et al. [83] performed a comparative study in rumor detection using different machine learning and deep learning models. The dataset used in this research was ArCOV19-Rumors. They used features such as n-grams, TF-IDF, and AraVec word embedding. The first experiment used SVM, KNN, NB, DT, RF, SGD, LR, and XGB for machine learning and evaluated the application of ensemble learning. In addition, six common deep learning models were used: Gated Recurrent Unit (GRU), RNN, LSTM, Bidirectional RNN (Bi-RNN), and Bi-GRU, BiLSTM with seven optimizers. The research showed that over-sampling did not enhance the performance of either the traditional or the deep learning models, and ensemble learning performed better than the single models. LSTM and Bi-LSTM with Root Mean Square Propagation (RMSprop) optimizer provided the best accuracy among the other deep learning models.

Al-Yahya et al. [45] performed another comparison study between the use of the NN model and the transformer-based language model for detecting Arabic fake news. The datasets used in this study included ArCOV19-Rumors, AraNews [84], Arabic News Stance (ANS) corpus [85], and COVID-19-Fakes [74]. The study used a linear model at word and character levels with Glove, Word2Vec, FastText, and document level. For the classification tasks, deep learning models CNN, GRU, and RNN were used. Moreover, QARiB, AraBERT, ArELectra, ARBERT, and MARBERT were used from the transformer-based models. The results showed that the transformer-based models performed better than the NN-based solutions, registering a 95% F1-score for QARiB. Fouad et al. [86] conducted another experiment for detecting fake news. The research used two datasets. The first contains news and tweets collected manually by the researchers and annotated to rumor or non-rumor, while the second dataset was from [87]. The two datasets were merged and used to create a third one. Word embedding and TensorFlow were used for text representation, and a set of traditional machine learning was used, which included Linear Support Vector Classifier (LSVC), SVC, MNB, BNB, SGD, DT, and RF. On the other hand, they examined a set of deep learning models, namely, CNN, LSTM, CNN+LSTM, BiLSTM, and CNN + BiLSTM. As a result, they found that no single model performed optimally over the three categories of datasets from the traditional machine learning models. For deep learning, the BiLSTM model provided the highest accuracy across all three datasets. Meanwhile, Shishah [88] performed another study, proposing a model called JointBERT for detecting the Arabic language. JointBERT in this research used Named Entity Recognition (NER) and Relative Features Classification (RFC) as parameters. The datasets used in this research were COVID-19-Fakes [74], ANS, Satirical [89], and AraNews. The results showed that JointBert outperformed the baseline results. The use of NER increased the performance because of its ability to extract news entities, which supports the model's performance in detecting fake news.

Some researchers have focused on specific types of misinformation, such as Bsoul et al. [90]. They built a dataset for clickbait detection, which facilitated automatic

[3] https://www.kaggle.com/c/fake-news/data?select=test.csv

classification and detection of news headlines. The study used BOW and TF-IDF and features related to headlines, such as headline length and with demonstrative pronouns, question words, and question mark. The researchers performed an experiment using SVM, LR, DT, NN. NB, SGD, and RF. These models produced a Macro F1-score of up to 0.81, which shows the effectiveness of using these seven machine learning models in the detection of clickbait news headlines. Research[89] focused in another type of fake news which is satire news. The research used dataset collected from different news websites and working to exploit textual features for the purpose of identifying satire news. For feature extraction they used count vectors, word-level TF-IDF, N-grams, Char-level, with MNB and XGB machine learning. In addition, the research used CNN with pre-trained word embedding which provided best accuracy registered which are 98.59%. The research found that satire news in Arabic language incline to have subjective tone with more positive and negative key terms. Research [85] release a dataset for claim verification, which are derived from subset of news title from Arabic News Text (ANT) corpus[94]. The authors modified news title to generate fake claims. For classification process, they used BERT, and LSTM for training and testing datasets. The study reported that BERT provides best F1-score registered which are 64.3%. Compared to this research, Nagoudi[84] used the same dataset from [85] in addition to their dataset which is automatically generated fake news from real one. The impact of this generated data on verification fake news are tested using transformer-based pre-trained models and compared with human created fake news dataset. The research reported that generated news are positively affect the fake news detection, and achieved better performance than [85] with 70% for F1-score.

In addition, [91] collected a dataset containing articles as fake and real, covering a single topic, which is Al Hajj. For the real articles, they collected articles in three dialects from Arab countries: Saudi, Egyptian, and Jordanian. The veracity of these articles was assessed using different fact-checking platforms. On the other hand, the study used crowdsourcing to create a fake news article based on a real one. The research did not provide a new classification method but suggested a set of features to provide an accurate Arabic classifier. The researchers built a lexical wordlist and an Arabic natural language processing (ANLP) architectural tool to extract the textual features, including POS, emotion, polarity, and syntactic. Based on these features, RF, SVM, and NB were used and tested for each single feature and a combination of features. The best result was registered for RF, with a combination of POS, syntactic and semantic roles, and contextual polarity features, which achieved an F1-score of 79%. Moreover, the research tested against human performance, providing 86% of articles classified correctly. Insufficient emphasis is being placed on the utilization of visual features in the realm of fake news detection, but recently, Albalawi et al. [52] proposed a model based on textual and image features. Their model consisted of three sub-models. For extracting the textual features, the BERT model was used. Meanwhile, two ensembles of pre-trained vision models were used for extracting the visual features (VGG-19 and ResNet50). The final model is a multimodal model used for concatenating the extracted textual and image features to represent a rumor vector. As a result, their proposed multimodal did not outperform the model with textual-based features. This shows that textual features are still considered pioneering in detecting rumors. Table II provides a brief summary of Arabic fake news detection studies.

V. Arabic Fact-Checking Websites

When researchers build datasets for fake news detection, they mostly rely on fact-checking websites. Fact-checking websites can be defined as platforms that evaluate the veracity of claims and information spread over social media networks, web articles, and public statements [95]. These kinds of websites usually rely on human experts who check the veracity of the news. Some fact-checking websites employ techniques that rely on evidence-based analysis or assess the credibility of the statements and how they align with the factual reality [95]. In the Arab countries, there are different fact-checking websites that researchers use as a starting point when building datasets for detecting fake news and rumors. It is worthy to provide researchers in the field with a list that can help them during their dataset-building phase. The common Arabic fact-checking websites will be described as follows:

- Norumors[4]: This is a standalone project that started in 2012 as a Twitter account searching for rumors and detecting their veracity. Thereafter, a website was established, which contains archives for different rumors and fake news that spread in Arab countries in general, especially in Saudi Arabia. The aim of this site is to spread awareness about rumors and expose the disseminators of falsehoods.

- Fatabyyano[5]: This is a standalone platform in the field of news fact-checking. First established in 2014 as a single page on Facebook. Fatabyyano uses the Facebook rating system, which has nine rating options, namely, false, partly false, true, false headline, satire, not eligible, opinion, prank generator, and not rated. Fatabyyano is certified by the International Fact-Checking Network (IFCN).

- Misbar[6]: Considered one of the leading fact-checking platforms in the Arab regions, it covers the Middle East and South African countries. The news classification system in Misbar using these classes includes fake, misleading, true, myth, selective, commotion, and satire.

- Akeed [7] : This is a fact-checking platform that is responsible for tracking the credibility of the Jordanian media. It was established with the support of the King Abdullah Fund for Development. The news are rated as false, biased, misleading, ambiguous, incomplete, inciting news, or contains an error.

[4] http://norumors.net/?post_type=rumors
[5] https://fatabyyano.net/
[6] https://misbar.com/
[7] https://akeed.jo/public/

- Verify[8]: This is considered the first Syrian fact-checking platform established in 2011 during the Syrian Revolution. In this platform, news are classified into three main classes depending on risk [red (high risk), orange (medium risk), and yellow (low risk)].

- Falso[9]: This is a Libyan platform for monitoring hate speech and a fact-checker in media news. It fact-checks both traditional news from TV and newspaper and also covers social media platforms and other websites.

- FactuelAFP Arabic[10]: This is the Arabic division of the French news agency dedicated to providing news and information. This division holds certification from IFCN and collaborates closely with Facebook to combat misinformation.

- Maharat-news fact-o-meter[11]: This is another fact-checking website certified by IFCN. Their main focus is detecting rumors on online media in Lebanon and other Arab regions. The rating system is based on three labels: true, partially true, and not true.

- Kashif[12]*:* it is an independent Palestinian fact-checking website that main goal to combat misleading information in Palestinian media. The rating system is based on nine labels: Manipulated, incorrect linking, outdated, sarcasm or parody, fabricated, false context, Impersonated and inaccurate content.

- Dabegad[13]: This is a fact-checking website in the Egyptian dialect that started in 2013 and aims to find and expose hoaxes in social media in Middle East countries.

VI. Arabic Fake News Datasets

It is clear that the first essential step to building an effective fake news detection system is constructing an appropriate dataset. There is still no agreed benchmark dataset for fake news detection in the Arabic language. Researchers in fake news detection usually focus on analyzing the text features only; however, this may no longer be enough, especially with the evolution of social media platforms and the diverse representation of fake news and their complexity. As such, different studies have recently used features related to news publishers, as well as social context features, such as user information and network information [75]. This was evident from the previous sections, which showed promising effectiveness. Accordingly, this section will briefly mention a set of publicly available datasets in the Arabic language. According to type, datasets can be classified into *social media posts* and *news articles*. In social media post datasets, in addition to the text content of posts, user and network information are utilized to detect fake news. Meanwhile, in news article datasets, researchers detect fake news by utilizing the headlines and the body of the articles.

One example of datasets related to posts in social media is [64], which included 9,000 tweets considered topic-independent. The dataset contains information related to the users and content, in addition to the sentiment features. The annotation process was performed manually, classifying the tweets into credible and noncredible. Alzanin and Azmi [67] constructed datasets that consisted of 27,100 posts related to 177 events from different domains, where 88 of these events were considered non-rumors, while 89 were rumors. This dataset also includes some features from users and contents. During the COVID-19 pandemic, researchers were encouraged to construct datasets covering this topic [74][76]. Furthermore, [74] constructed a bilingual dataset that covers the Arabic and English languages. The dataset was collected from Twitter using keywords related to COVID-19. The dataset has been automatically annotated using 13 machine learning classifiers and seven feature extraction techniques, including TF, TF-IDF (n-gram, character level), and word embedding. The dataset has two labels: real and misleading. The tweets' IDs and detection labels are made publicly available to other researchers by the authors.

AraCOVID19-MFH [76] is a multilabel dataset manually annotated to 10 labels. It is an abbreviation of the Arabic COVID-19 Multilabel Fake News and Hate Speech Detection dataset. The dataset consists of 10,828 items of annotated data that include MSA and different dialects. The dataset was collected by searching some keywords from Twitter. This dataset can be used for both hate speech and fake news detection tasks. In [9], another dataset was collected from Twitter by searching for a specific list of keywords. The dataset is labeled as misleading (1311 tweets) and others (7475 tweets). It consists of 8,786 tweets in total, and it is clear that this dataset suffers from imbalance. The researchers only published the ID of the tweets and their corresponding labels (misleading and other). Moreover, in [73], the dataset was collected from Twitter based on the COVID-19 domain. Part of the dataset was annotated manually first and then used to train different machine learning models to automatically produce annotation for the rest of the unlabeled data. They were annotated as fake and genuine and only relied on the text feature. The authors published the tweet's text and its label without the ID. ArCorona [96] is a larger dataset collected manually from Twitter in the health domain during the early stages of COVID-19. The dataset contains 30 million tweets, with 8,000 tweets labeled into 13 labels. The dataset contains different dialects from the Arabic regions.

Alsudais and Rayson [62] also manually collected a dataset that contains one million tweets about COVID-19, among which 2,000 were annotated to 895 false, 0 unrelated, and 789 true. The tweets were collected using keywords related to infectious diseases. All the above-mentioned datasets rely on textual content; however, there is a dataset called ArCOV19-Rumors [75], which consists of 9,414 tweets that were manually labeled as false (1,753), true (1,831), and others (5,830), which are related to 138 claims This dataset works on two levels: claim-level verification and tweet-level verification, both of which have two labels, namely, true and false. and their correspond relevant tweets. Meanwhile, tweet-level verification contains the tweet and the propagation

[8] https://verify-sy.com/
[9] https://falso.ly/
[10] https://factcheck.afp.com/ar/list
[11] https://maharat-news.com/fact-o-meter
[12] https://kashif.ps/category/facts-in-en/
[13] https://dabegad.com/

network (retweets, replies). Moreover, study in [82] built dataset about Covid-19. The datasets are labeled for binary classification and multiclass, based on seven questions. The datasets available in four language which is English, Arabic, Dutch, and Bulgarian. The main focus for this paper is fake news, so 4966 Arabic tweets are classified into (815) contains fake information and (2602) not contains false information.

TABLE III. Summary of Social Media Datasets for Arabic Fake News Detection. ‘E’ Evidence, ‘T’ Textual, ‘V’ Visual, ‘U’ User, ‘P’ Post, ‘N’ Network, ‘M’ Manually, ‘A’ Automatically

Dataset	Domain	Size	Annotated	Annotation type		Labels	Features used						Source
				M	*A*		*E*	*T*	*V*	*U*	*P*	*N*	
Dataset [64]	Multi-domain	9K tweets	9K tweets	✓	-	(5400) credible (3600) noncredible	-	✓	-	✓	✓	-	Twitter
Dataset [67]	Multi-domain	271K tweets 177 events	271K tweets	✓	-	(88) events non-rumor (89) events rumor	-	✓	-	✓	✓	-	
COVID-19-FAKE [74]	Health (covid-19)	220K tweets	220K tweets	-	✓	Misleading Real	-	✓	-	-	-	-	
AraCOVID19-MFH[76]	Health (covid-19)	300K tweets	10828 tweets	✓	-	10 different labels	-	✓	-	-	-	-	
Dataset [9]	Health (covid-19)	4.5M tweets	8.8K tweets	✓	-	(1,311) misleading, (7,475) other	-	✓	-	-	-	-	
Dataset [73]	Health (covid-19)	36066 tweets	36066 tweets	✓	✓	(20417) Fake (15649) Not fake	-	✓	-	-	-	-	
ArCorona[96]	Health (covid-19)	30M tweets	8K tweets	✓	-	13 different labels	-	✓	-	-	-	-	
Dataset [62]	Health (covid-19)	1M tweets	2K tweets	✓	-	(316) False (895) True (789) Unrelated	-	✓	-	-	-	-	
ArCovid19-Rumors[75]	Health (covid-19)	1M tweets 138 claims	9414 tweets	✓	-	(1753) false (1831) true (5830) other	-	✓	-	✓	✓	✓	
Dataset [82]	Health (covid-19)	4966 tweets	4966 tweets	✓	-	(2609) No (815) Yes	-	✓	-	-	-	-	
Dataset [7]	Politic	3358 tweets	2708 tweets	✓	-	(1570) credible (1138) non-credible	-	✓	-	-	-	-	

TABLE IV. Summary of News Article Datasets for Arabic Fake News Detection. ‘E’ Evidence, ‘T’ Textual, ‘V’ Visual, ‘S’ Stance

Dataset	Domain	Size	Annotation type	Labels	Features				Source
					E	*T*	*V*	*S*	
AraNews [84]	Multi-domain	5187957 news articles	Manually	False, True	-	✓	-	-	50 newspapers.
Dataset[72]	Multi-domain	323 articles		(100) reliable (223) unreliable	-	✓	-	-	Kashif fact-checking website, social media, WhatsApp group, and news site.
AraFact [97]	Multi-domain	6222 claims		(4037) false (1891) partly-false (198) True (90) Sarcasm (6) unverifiable	✓	✓	✓	-	5 Arabic fact-checking websites
Dataset [98]	Politic (Syria War)	422 claims 3042documents		(219) false claim. (203) true claim Documents: (1239) false (1803) True	-	✓	-	✓	2 websites VERIFY for false claim and REUTERS[14] for True claim
ANS [85]	Multi-domain	4547 claims (3786) pairs (claim, evidence)		(3072) True (1475) False	-	✓	-	✓	News headlines from media sources in Middle east and ANT corpus.
AraStance [99]	Multi-domain	910 claims 4063 pair (claim, articles)		(606) False claim (304) True claims Articles: (2421) False (1642) True	-	✓	-	✓	3 fact-checking websites Aranews, Dabegad, Norumors, REUTERS
Satirical fake News[89]	Politic	3185 articles		(3185) fake. (3710) real.	-	✓	-	-	2 satirical news websites for fake news. And Official news site for real.
AFND [100]	Multi-domain	606912 articles		(207310) credible (167233) not credible (232369) undecided	-	✓	-	-	134 Arabic online news sources
Dataset. [7]	Multi-domain	175 blog posts		(100) credible (57) fairly credible (18) non-credible	-	✓	-	-	Arabic news articles

[14] http://ara.reuters.com

For the news article dataset, AraNews [84] is a large dataset collected from different countries, covering different topics. The dataset relied on a list of 50 newspapers from 15 Arab countries, the United Kingdom, and the United States. Based on this list, 5,187,957 news articles were collected and labeled as true or false. Also, research [72] published dataset which contains 323 articles from different domains. The dataset was labeled manually from two experts into reliable and unreliable. Different sources are used for collecting this dataset such as Kashif fact-checking websites, social media, and news sites. Ali et al. [97] produced a dataset called AraFacts that contains claims collected from five Arabic fact-checking websites. The dataset comprises URLs for fact-checking articles and links to evidence pages sourced from various outlets, enabling the extraction of images and supporting evidence. This making them the first to collect datasets in Arabic that combined texts, image contents, and evidence. Research in [52] used AraFacts to extract images and text. The dataset contained 6,222 claims annotated by professional fact-checkers manually.

Typically, researchers focus independently on specific tasks such as stance detection, fact-checking, document retrieval, and source credibility, but study [98] making these tasks available to be integrated using unified corpus. The study created a corpus consisted of 442 claims that covered topics related to the Syrian war and some political issues in the Middle East. Each claim was labeled as false or true based on factuality. In addition, 2,042 articles retrieved for these claims were annotated based on their stance into agree, disagree, discuss, and unrelated. In addition, the dataset included a rationale attribute that enabled the fact-checking system to provide explanations for its decisions. This attribute encompassed the display of extracted sentences or phrases from the retrieved documents, which served as illustrative examples of the detected stance. Similarly, ANS [85] is a dataset of news titles that were paraphrased and altered. ANS covers several topics collected from online news sites. The dataset contains two versions. Based on claims verification, 4,547 records were labeled fake and not fake; the other version consisted of claims and evidence pairs containing 3,786 records. The difference between ANS and Baly et al. [98] is that ANS generated fake and true claims from true news. In the same vein, the AraStance dataset [99] contained 4,063 pairs of claims and articles from multiple countries, covering topics in politics, health, sports, and others. The annotation process was performed manually according to the veracity of the claims, whether they are true or false, and according to the article's stance on the claims, which ended with four labels: agree, disagree, discuss, and unrelated. Satirical News [89] is a hand-built dataset that includes fake articles. The fake news articles were collected from two satirical news websites: Al-Hudood [15] and Al-Ahram Al-Mexici[16]. For the real news dataset, the research used an open-source datasets[17]. Arabic Fake News Dataset (AFND) [100] is a collection of news stories in Arabic that are available to the public and were gathered from Arabic news websites. A dataset used for detecting article credibility, it consisted of 606,912 articles labeled as credible (207,310), not credible (167,233), and undecided (232,369). The researchers used the Misbar fact-checking platform for classifying the articles into these three classes. Some research built datasets that contained both social media posts and news articles, such as Al Zaatari et al.'s dataset [7], which consisted of blog posts and tweets related to the Syrian crisis. This kind of corpora's main goal is to analyze the credibility of the news. It consisted of 2,708 tweets and 175 blog posts; the datasets, in general, are labeled as credible and not credible. Tables III and IV provide a summary of the social media posts and news article datasets. In the context of news article datasets, the range of available features is comparatively narrower compared to those accessible for social media datasets. Thus, our focus was confined to textual, visual, stance, and evidence features for the purpose of comparison.

VII. Discussion

Based on the previous sections, fake news detection in the Arabic language is still in its nascent stages compared with other languages. Although a multitude of efforts have been exerted in Arabic fake news detection, the process still suffers from various limitations. We can categorize them into limitations related to datasets, feature extraction, and classification algorithms based on the literature.

A. Datasets

There is no benchmark dataset in the Arabic language, and most of the datasets used in previous studies are not available online. Researchers need to enrich the fake news detection field by making their data available using any platform, such as their own page in GitHub[18] or MASADER[19] repository. Publicly available datasets enable other studies to exert robust efforts, and their results can be compared with others using the same datasets. This is one of the important factors to measure the improvement of performance among different studies. Moreover, the processes for collecting and preparing datasets are not always mentioned clearly in the research papers. Researchers in this field need accurate guidelines to undertake this process, for example, by providing a list of common sources for extracting appropriate news, the annotation process, the cleaning and preprocessing phase, and so on. In addition, most of the datasets suffer from unbalanced classes, where the real news category is usually larger than the false one such as [9] [64]. Moreover, when screening parts of these datasets, there were numerous instances for the same news, especially those collected from social media platforms such as Twitter. This repetition in the values of the news texts can be attributed to the fact that the same news can propagate among such platforms, and because the collection process usually depends on the prepared list of keywords, the probability of having the same news with the same exact text is higher. In some situations, having duplicated contents will decrease the performance of the classifier [101]. Another problem related to the datasets is the domain they cover. For example, Tables III and IV show that most of the datasets focused on covering the COVID-19 pandemic, while the others focused on whether

[15] https://alhudood.net/

[16] https://alahraam.com/

[17] https://sourceforge.net/projects/ar-text-mining/files/Arabic-Corpora/

[18] https://github.com/

[19] https://arbml.github.io/masader/

the data was a post from social media or articles covering different domains with unbalanced categories for each. These researchers argued that they adapted this process to find common features across various domains. However, they ignored the fact that each sector has its own features, which can help in detecting fake news related to them. Having datasets that cover a specific domain is important. Specialized terminology is prevalent within specific topic domains, and possessing technical expertise is essential for discerning the authenticity of each news item [18]. Hence, it is crucial to construct datasets tailored to the unique requirements of a particular topic domain, such as the SCIFACT dataset [102].

Moreover, Arabic countries cover different regions with different dialects; thus, datasets covering MSA and dialects require a complex process to mitigate this problem. We still need an Arabic dataset that covers the MSA language and others that cover different dialects. The dataset publishers usually release only the IDs for the tweets in accordance with Twitter policy. When other researchers try to extract the same tweets, they are already deleted or protected by their owners. This situation decreases the number of available datasets and is one of the immense problems that need to be solved by applying a repository that does not contain critical features and saves only appropriate and valuable ones. Working to extract more records to balance the dataset affected by this deletion is time-consuming. The creation of an unaltered dataset for the purpose of detecting fake news on social media platforms represents a crucial milestone in establishing a benchmark dataset. This endeavor aids in effectively assessing the performance of models and their ability to verify the accuracy of information with the collaboration of social media platforms.

B. Feature Extraction and Classification Process

Most previous studies have focused on extracting textual features in articles and social media posts, which is not usually sufficient to detect fake news [18]. Rather, it is important to combine features such as those related to users, posts, networks for social posts, and the metadata related to the articles, including the publisher's name, URLs, headlines, body, and comments. Based on the literature review, only a few researches are using a combination of these features. In addition, when looking for the visual contents, images provide an improvement in the performance of the classifier in other languages, which need more investigation in Arabic [50]. There are a few research in Arabic with this attribute, such as [52]. Extracting features from the replies of users is also an important aspect [56]. The only publicly available Arabic dataset that has considered users' responses in their dataset is ArCOVID19-Rumors. Employing sentiment analysis, stance detection, and emotion recognition with fake news detection still needs more investigation, especially in relation to the responses of the crowd. Moreover, features related to networks are still not investigated in fake news in the Arabic language, except for the work performed by [75]. Traditional machine learning approaches, such as SVM, LR, and KNN, are the most used in Arabic research. Based on the review, most research in Arabic has used language models, such as AraBERT, MARBERT, mBERT, and QARiB. Research using the ensemble techniques is scarce, and there is a need for more studies applying this model. Applying the ensemble methods using traditional machine learning and deep learning is another window of opportunity that needs to be investigated for fake news detection because ensemble methods have the ability to improve prediction performance with regard to some characteristics, such as overfitting avoidance, computational advantages, and representation [103].

VIII. Conclusion

This review revealed that a few studies related to detecting fake news have been conducted in the Arabic language, indicating that Arab researchers should allocate more attention to this issue. Most Arabic studies have focused on social media platforms, particularly Twitter. Most Arabic researchers have investigated events related to politics (e.g., Syrian crisis) and health (e.g., COVID-19 pandemic) and have created their own dataset for testing the proposed models. This may not be an effective approach because it results in many different datasets with a range of accuracy measurement results. A benchmark dataset that contains as many features as possible to help detect fake news should be used. Our future objectives encompass the development of an Arabic dataset encompassing diverse extraction features, including visual attributes, enabling us to explore their influence when combined with other features for the purpose of detecting fake news on social media.

References

[1] Kepios, “Global Social Media Statistics,” Datareportal, 2022. https://datareportal.com/social-media-users (accessed Nov. 10, 2023).

[2] M. Walker and K. E. Matsa, “News Consumption Across Social Media in 2021,” Pew Research Center, 2021. https://www.pewresearch.org/journalism/2021/09/20/news-consumption-across-social-media-in-2021/ (accessed Feb. 11, 2023).

[3] S. Vosoughi, D. Roy, and S. Aral, “The spread of true and false news online,” Science (80-.)., vol. 359, no. 6380, pp. 1146–1151, 2018, doi: 10.1126/science.aap9559.

[4] A. Al-Rawi, A. Fakida, and K. Grounds, “Investigation of COVID-19 Misinformation in Arabic on Twitter: Content Analysis,” JMIR Infodemiology, vol. 2, no. 2, p. e37007, Jul. 2022, doi: 10.2196/37007.

[5] S. Kula and R. K. P. W. M. Choraś Michałand Kozik, “Sentiment Analysis for Fake News Detection by Means of Neural Networks,” in Computational Science -- ICCS 2020, 2020, pp. 653–666.

[6] P. Nakov, F. Alam, S. Shaar, G. Martino, and Y. Zhang, “A Second Pandemic? Analysis of Fake News About COVID-19 Vaccines in Qatar,” ArXiv, vol. abs/2109.1, 2021.

[7] A. Al Zaatari et al., “Arabic Corpora for Credibility Analysis,” in Proceedings of the Tenth International Conference on Language Resources and Evaluation ({LREC}’16), May 2016, pp. 4396–4401, [Online]. Available: https://aclanthology.org/L16-1696.

[8] M. LIPKA and C. HACKETT, “Why Muslims are the world’s fastest-growing religious group,” Pew Research Center, 2017. https://www.pewresearch.org/short-reads/2017/04/06/why-muslims-are-the-worlds-fastest-growing-religious-group/ (accessed Nov. 10, 2023).

[9] S. Alqurashi, B. Hamoui, A. S. Alashaikh, A. Alhindi, and E. A. Alanazi, “Eating Garlic Prevents COVID-19 Infection: Detecting Misinformation on the Arabic Content of Twitter,” ArXiv, vol. abs/2101.0, 2021.

[10] H. ALSaif and T. Alotaibi, “Arabic Text Classification using Feature-Reduction Techniques for Detecting Violence on Social Media,” Int. J. Adv. Comput. Sci. Appl., vol. 10, May 2019, doi: 10.14569/IJACSA.2019.0100409.

[11] S. K. Hamed, M. J. Ab Aziz, and M. R. Yaakub, “A review of fake news detection approaches: A critical analysis of relevant studies and

highlighting key challenges associated with the dataset, feature representation, and data fusion.," Heliyon, vol. 9, no. 10, p. e20382, Oct. 2023, doi: 10.1016/j.heliyon.2023.e20382.

[12] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake News Detection on Social Media: A Data Mining Perspective," SIGKDD Explor. Newsl., vol. 19, no. 1, pp. 22–36, Sep. 2017, doi: 10.1145/3137597.3137600.

[13] R. Mouty and A. Gazdar, "Survey on Steps of Truth Detection on Arabic Tweets," in 2018 21st Saudi Computer Society National Computer Conference (NCC), 2018, pp. 1–6, doi: 10.1109/NCG.2018.8593060.

[14] S. Althabiti, M. Alsalka, and E. Atwell, "A Survey: Datasets and Methods for Arabic Fake News Detection," Int. J. Islam. Appl. Comput. Sci. Technol., vol. 11, pp. 19–28, 2023.

[15] R. A. M. San Ahmed, "A Novel Taxonomy for Arabic Fake News Datasets," Int. J. Comput. Digit. Syst., vol. 14, no. 1, pp. 159–166, 2023, doi: 10.12785/ijcds/140115.

[16] Z. I. Mahid, S. Manickam, and S. Karuppayah, "Fake News on Social Media: Brief Review on Detection Techniques," in 2018 Fourth International Conference on Advances in Computing, Communication Automation (ICACCA), 2018, pp. 1–5, doi: 10.1109/ICACCAF.2018.8776689.

[17] A. D'Ulizia, M. C. Caschera, F. Ferri, and P. Grifoni, "Fake news detection: a survey of evaluation datasets," PeerJ. Comput. Sci., vol. 7, pp. e518–e518, Jun. 2021, doi: 10.7717/peerj-cs.518.

[18] T. Murayama, "Dataset of Fake News Detection and Fact Verification: A Survey," arXiv, 2021, [Online]. Available: http://arxiv.org/abs/2111.03299.

[19] S. Kumar, R. West, and J. Leskovec, "Disinformation on the Web: Impact, Characteristics, and Detection of Wikipedia Hoaxes," in Proceedings of the 25th International Conference on World Wide Web, 2016, pp. 591–602, doi: 10.1145/2872427.2883085.

[20] V. Rubin, Y. Chen, and N. Conroy, "Deception detection for news: Three types of fakes," Proc. Assoc. Inf. Sci. Technol., vol. 52, pp. 1–4, 2015, doi: 10.1002/pra2.2015.145052010083.

[21] C. Buntain and J. Golbeck, "Automatically Identifying Fake News in Popular Twitter Threads," in 2017 IEEE International Conference on Smart Cloud (SmartCloud), 2017, pp. 208–215, doi: 10.1109/SmartCloud.2017.40.

[22] R. Ghanem and H. Erbay, "Context-dependent model for spam detection on social networks," SN Appl. Sci., vol. 2, no. 9, p. 1587, 2020, doi: 10.1007/s42452-020-03374-x.

[23] P. Hernon, "Disinformation and misinformation through the internet: Findings of an exploratory study," Gov. Inf. Q., vol. 12, no. 2, pp. 133–139, 1995, doi: https://doi.org/10.1016/0740-624X(95)90052-7.

[24] S. Volkova, K. Shaffer, J. Y. Jang, and N. Hodas, "Separating Facts from Fiction: Linguistic Models to Classify Suspicious and Trusted News Posts on Twitter," in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Jul. 2017, pp. 647–653, doi: 10.18653/v1/P17-2102.

[25] L. Wu, F. Morstatter, K. M. Carley, and H. Liu, "Misinformation in Social Media: Definition, Manipulation, and Detection," SIGKDD Explor. Newsl., vol. 21, no. 2, pp. 80–90, Nov. 2019, doi: 10.1145/3373464.3373475.

[26] Y. Chen, N. J. Conroy, and V. L. Rubin, "Misleading Online Content: Recognizing Clickbait as 'False News,'" in Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection, 2015, pp. 15–19, doi: 10.1145/2823465.2823467.

[27] J. Brummette, M. DiStaso, M. Vafeiadis, and M. Messner, "Read All About It: The Politicization of 'Fake News' on Twitter," Journal. \& Mass Commun. Q., vol. 95, no. 2, pp. 497–517, 2018, doi: 10.1177/1077699018769906.

[28] C. Burfoot and T. Baldwin, "Automatic Satire Detection: Are You Having a Laugh?," in Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, Aug. 2009, pp. 161–164, [Online]. Available: https://aclanthology.org/P09-2041.

[29] C. Lumezanu, N. Feamster, and H. Klein, "#bias: Measuring the Tweeting Behavior of Propagandists," Proc. Int. AAAI Conf. Web Soc. Media, vol. 6, no. 1, pp. 210–217, 2021, [Online]. Available: https://ojs.aaai.org/index.php/ICWSM/article/view/14247.

[30] G. S.Jowett and V. O'Donnell, Propaganda & persuasion. SAGE, 2014.

[31] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, and B. Stein, "A Stylometric Inquiry into Hyperpartisan and Fake News," in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Jul. 2018, pp. 231–240, doi: 10.18653/v1/P18-1022.

[32] S. Zannettou, M. Sirivianos, J. Blackburn, and N. Kourtellis, "The Web of False Information: Rumors, Fake News, Hoaxes, Clickbait, and Various Other Shenanigans," J. Data Inf. Qual., vol. 11, no. 3, May 2019, doi: 10.1145/3309699.

[33] Amy Watson, "Fake news worldwide - Statistics & Facts," 2020. https://www.statista.com/topics/6341/fake-news-worldwide/#dossierKeyfigures (accessed Nov. 10, 2021).

[34] A. Waston, "Fake news in the U.S. - statistics & facts | Statista," Jun. 16, 2021. https://www.statista.com/topics/3251/fake-news/ (accessed Nov. 10, 2021).

[35] G. Rannard, "Australia fires : Misleading maps and pictures go viral," BBC Trending, 2020. https://www.bbc.com/news/blogs-trending-51020564.

[36] M. Sallam et al., "High Rates of COVID-19 Vaccine Hesitancy and Its Association with Conspiracy Beliefs: A Study in Jordan and Kuwait among Other Arab Countries," Vaccines, vol. 9, 2021.

[37] J. Klausen, "Tweeting the Jihad: Social Media Networks of Western Foreign Fighters in Syria and Iraq," Stud. Confl. Terror., vol. 38, no. 1, pp. 1–22, Jan. 2015, doi: 10.1080/1057610X.2014.974948.

[38] M. A. Alonso, D. Vilares, C. Gómez-Rodríguez, and J. Vilares, "Sentiment Analysis for Fake News Detection," Electronics, vol. 10, p. 1348, 2021.

[39] J. Z. Pan, S. Pavlova, C. Li, N. Li, Y. Li, and J. Liu, "Content Based Fake News Detection Using Knowledge Graphs BT - The Semantic Web – ISWC 2018," 2018, pp. 669–683.

[40] J. Dougrez-Lewis, E. Kochkina, M. Arana-Catania, M. Liakata, and Y. He, "PHEMEPlus: Enriching Social Media Rumour Verification with External Evidence," in Proceedings of the Fifth Fact Extraction and VERification Workshop (FEVER), May 2022, pp. 49–58, doi: 10.18653/v1/2022.fever-1.6.

[41] K. Popat, S. Mukherjee, A. Yates, and G. Weikum, "DeClarE: Debunking Fake News and False Claims using Evidence-Aware Deep Learning," in Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018, pp. 22–32, doi: 10.18653/v1/D18-1003.

[42] A. Bondielli and F. Marcelloni, "A Survey on Fake News and Rumour Detection Techniques," Inf. Sci., vol. 497, no. C, pp. 38–55, Sep. 2019, doi: 10.1016/j.ins.2019.05.035.

[43] C. Castillo, M. Mendoza, and B. Poblete, "Information Credibility on Twitter," in Proceedings of the 20th International Conference on World Wide Web, 2011, pp. 675–684, doi: 10.1145/1963405.1963500.

[44] H. Ahmed, I. Traoré, and S. Saad, "Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques," 2017.

[45] M. Al-Yahya, H. Al-Khalifa, H. Al-Baity, D. AlSaeed, and A. Essam, "Arabic Fake News Detection: Comparative Study of Neural Networks and Transformer-Based Approaches," Complexity, vol. 2021, p. 5516945, 2021, doi: 10.1155/2021/5516945.

[46] S. Kwon, M. Cha, and K. Jung, "Rumor Detection over Varying Time Windows," PLoS One, vol. 12, no. 1, pp. 1–19, 2017, doi: 10.1371/journal.pone.0168344.

[47] N. Ruchansky, S. Seo, and Y. Liu, "CSI: A Hybrid Deep Model for Fake News Detection," Proc. 2017 ACM Conf. Inf. Knowl. Manag., 2017.

[48] Y. Liu and Y.-F. Wu, "Early Detection of Fake News on Social Media Through Propagation Path Classification with Recurrent and Convolutional Networks," 2018.

[49] Z. Jin, J. Cao, Y. Zhang, J. Zhou, and Q. Tian, "Novel Visual and Statistical Image Features for Microblogs News Verification," IEEE Trans. Multimed., vol. 19, no. 3, pp. 598–608, 2017, doi: 10.1109/TMM.2016.2617078.

[50] S. K. Uppada, P. Patel, and S. B., "An image and text-based multimodal model for detecting fake news in OSN's," J. Intell. Inf. Syst., 2022, doi: 10.1007/s10844-022-00764-y.

[51] Y. Wang et al., “EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection,” in Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery &amp; Data Mining, 2018, pp. 849–857, doi: 10.1145/3219819.3219903.

[52] R. M. Albalawi, A. T. Jamal, A. O. Khadidos, and A. M. Alhothali, “Multimodal Arabic Rumors Detection,” IEEE Access, vol. 11, pp. 9716–9730, 2023, doi: 10.1109/ACCESS.2023.3240373.

[53] Z. S. Ali, A. Al-Ali, and T. Elsayed, “Detecting Users Prone to Spread Fake News on Arabic Twitter,” in Proceedinsg of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools with Shared Tasks on Qur’an QA and Fine-Grained Hate Speech Detection, 2022, pp. 12–22.

[54] R. Alghamdi and O. Alrwais, “Towards Automatic Rumor Detection in Arabic Tweets,” Int. J. Data Min. Manag. Syst., vol. 1, no. 5, pp. 1–14, 2022.

[55] G. Jardaneh, H. Abdelhaq, M. Buzz, and D. Johnson, “Classifying Arabic tweets based on credibility using content and user features,” in 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology, JEEIT 2019 - Proceedings, 2019, pp. 596–601, doi: 10.1109/JEEIT.2019.8717386.

[56] H. Kidu, H. Misgna, T. Li, and Z. Yang, “User Response-Based Fake News Detection on Social Media BT - Applied Informatics,” 2021, pp. 173–187.

[57] E. Tacchini, G. Ballarin, M. L. Della Vedova, S. Moret, and L. de Alfaro, “Some Like it Hoax: Automated Fake News Detection in Social Networks,” ArXiv, vol. abs/1704.0, 2017.

[58] M. Davoudi, M. R. Moosavi, and M. H. Sadreddini, “DSS: A hybrid deep model for fake news detection using propagation tree and stance network,” Expert Syst. Appl., vol. 198, p. 116635, 2022, doi: https://doi.org/10.1016/j.eswa.2022.116635.

[59] N. Zhong, G. Zhou, W. Ding, and J. Zhang, “A Rumor Detection Method Based on Multimodal Feature Fusion by a Joining Aggregation Structure,” Electronics, vol. 11, no. 19, 2022, doi: 10.3390/electronics11193200.

[60] S. Kwon, M. Cha, K. Jung, W. Chen, and Y. Wang, “Prominent Features of Rumor Propagation in Online Social Media,” in 2013 IEEE 13th International Conference on Data Mining, 2013, pp. 1103–1108, doi: 10.1109/ICDM.2013.61.

[61] F. L. Alotaibi and M. M. Alhammad, “Using a Rule-based Model to Detect Arabic Fake News Propagation during Covid-19,” Int. J. Adv. Comput. Sci. Appl., vol. 13, no. 1, 2022, doi: 10.14569/IJACSA.2022.0130114.

[62] L. Alsudias and P. Rayson, “COVID-19 and Arabic Twitter: How can Arab World Governments and Public Health Organizations Learn from Social Media?,” Jul. 2020, [Online]. Available: https://aclanthology.org/2020.nlpcovid19-acl.16.

[63] M. Alkhair, K. Meftouh, K. Smaïli, and N. Othman, “An Arabic Corpus of Fake News: Collection, Analysis and Classification,” in Arabic Language Processing: From Theory to Practice, 2019, pp. 292–302.

[64] R. El Ballouli, W. El-Hajj, A. Ghandour, S. Elbassuoni, H. Hajj, and K. Shaban, “CAT: Credibility Analysis of Arabic Content on Twitter,” in WANLP 2017, co-located with EACL 2017 - 3rd Arabic Natural Language Processing Workshop, Proceedings of the Workshop, Apr. 2017, pp. 62–71, doi: 10.18653/v1/w17-1308.

[65] R. Mouty and A. Gazdar, “Employing the Google Search and Google Translate to Increase the Performance of the Credibility Detection in Arabic Tweets BT - Computational Collective Intelligence,” 2022, pp. 781–788.

[66] T. Thaher, M. Saheb, H. Turabieh, and H. Chantar, “Intelligent Detection of False Information in Arabic Tweets Utilizing Hybrid Harris Hawks Based Feature Selection and Machine Learning Models,” Symmetry (Basel)., vol. 13, no. 4, 2021, doi: 10.3390/sym13040556.

[67] S. M. Alzanin and A. M. Azmi, “Rumor Detection in Arabic Tweets Using Semi-Supervised and Unsupervised Expectation–Maximization,” Know.-Based Syst., vol. 185, no. C, Dec. 2019, doi: 10.1016/j.knosys.2019.104945.

[68] F. Saeed, W. M.S., M. Al-Sarem, and E. Abdullah, “Detecting Health-Related Rumors on Twitter using Machine Learning Methods,” Int. J. Adv. Comput. Sci. Appl., vol. 11, Jan. 2020, doi: 10.14569/IJACSA.2020.0110842.

[69] S. Alyoubi, M. Kalkatawi, and F. Abukhodair, “The Detection of Fake News in Arabic Tweets Using Deep Learning,” Appl. Sci., vol. 13, no. 14, 2023, doi: 10.3390/app13148209.

[70] M. Abdul-Mageed, A. Elmadany, and E. M. B. Nagoudi, “ARBERT & MARBERT: Deep Bidirectional Transformers for Arabic,” ArXiv, vol. abs/2101.0, 2020.

[71] H. M. Alawadh, A. Alabrah, T. Meraj, and H. T. Rauf, “Attention-Enriched Mini-BERT Fake News Analyzer Using the Arabic Language,” Futur. Internet, vol. 15, no. 2, 2023, doi: 10.3390/fi15020044.

[72] R. Assaf and M. Saheb, “Dataset for Arabic Fake News,” in 2021 IEEE 15th International Conference on Application of Information and Communication Technologies (AICT), 2021, pp. 1–4, doi: 10.1109/AICT52784.2021.9620228.

[73] A. Mahlous and A. Al-Laith, “Fake News Detection in Arabic Tweets during the COVID-19 Pandemic,” Int. J. Adv. Comput. Sci. Appl., vol. 12, 2021, doi: 10.14569/IJACSA.2021.0120691.

[74] M. K. Elhadad, K. F. Li, and F. Gebali, “COVID-19-FAKES: A Twitter (Arabic/English) Dataset for Detecting Misleading Information on COVID-19,” 2020.

[75] F. Haouari, M. Hasanain, R. Suwaileh, and T. Elsayed, “ArCOV19-Rumors: Arabic COVID-19 Twitter Dataset for Misinformation Detection,” in Proceedings of the Sixth Arabic Natural Language Processing Workshop, Apr. 2021, pp. 72–81, [Online]. Available: https://aclanthology.org/2021.wanlp-1.8.

[76] M. S. Hadj Ameur and H. Aliane, “AraCOVID19-MFH: Arabic COVID-19 Multi-label Fake News & Hate Speech Detection Dataset,” Procedia Comput. Sci., vol. 189, pp. 232–241, 2021, doi: https://doi.org/10.1016/j.procs.2021.05.086.

[77] A. B. Nassif, A. Elnagar, O. Elgendy, and Y. Afadar, “Arabic fake news detection based on deep contextualized embedding models,” Neural Comput. Appl., vol. 34, no. 18, pp. 16019–16032, 2022, doi: 10.1007/s00521-022-07206-4.

[78] “Fake News | Kaggle,” 2018. https://www.kaggle.com/c/fake-news/data?select=test.csv (accessed Feb. 06, 2023).

[79] I. Touahri and A. Mazroui, “EvolutionTeam at CLEF2020-CheckThat! lab: Integration of Linguistic and Sentimental Features in a Fake News Detection Approach.,” 2020.

[80] M. Abd Elaziz, A. Dahou, D. A. Orabi, S. Alshathri, E. M. Soliman, and A. A. Ewees, “A Hybrid Multitask Learning Framework with a Fire Hawk Optimizer for Arabic Fake News Detection,” Mathematics, vol. 11, no. 2, 2023, doi: 10.3390/math11020258.

[81] F. Husain, “OSACT4 Shared Task on Offensive Language Detection: Intensive Preprocessing-Based Approach,” in Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection, May 2020, pp. 53–60, [Online]. Available: https://aclanthology.org/2020.osact-1.8.

[82] F. Alam et al., “Fighting the COVID-19 Infodemic: Modeling the Perspective of Journalists, Fact-Checkers, Social Media Platforms, Policy Makers, and the Society,” in Findings of the Association for Computational Linguistics: EMNLP 2021, Nov. 2021, pp. 611–649, doi: 10.18653/v1/2021.findings-emnlp.56.

[83] G. Amoudi, R. Albalawi, F. Baothman, A. Jamal, H. Alghamdi, and A. Alhothali, “Arabic rumor detection: A comparative study,” Alexandria Eng. J., vol. 61, no. 12, pp. 12511–12523, 2022, doi: https://doi.org/10.1016/j.aej.2022.05.029.

[84] E. M. B. Nagoudi, A. A. Elmadany, M. Abdul-Mageed, T. Alhindi, and H. Cavusoglu, “Machine Generation and Detection of Arabic Manipulated and Fake News,” CoRR, vol. abs/2011.0, 2020, [Online]. Available: https://arxiv.org/abs/2011.03092.

[85] J. Khouja, “Stance Prediction and Claim Verification: An Arabic Perspective,” in Proceedings of the Third Workshop on Fact Extraction and VERification (FEVER), Jul. 2020, pp. 8–17, doi: 10.18653/v1/2020.fever-1.2.

[86] K. M. Fouad, S. F. Sabbeh, and W. Medhat, “Arabic Fake News Detection Using Deep Learning,” Comput. Mater. \& Contin., vol. 71, no. 2, pp. 3647–3665, 2022, doi: 10.32604/cmc.2022.021449.

[87] F. Rangel, P. Rosso, A. Charfi, W. Zaghouani, B. Ghanem, and J. Snchez-Junquera, “Overview of the track on author profiling and deception detection in arabic,” 2019.

[88] W. Shishah, “JointBert for Detecting Arabic Fake News,” IEEE Access, vol. 10, pp. 71951–71960, 2022, doi: 10.1109/ACCESS.2022.3185083.

[89] H. Saadany, C. Orasan, and E. Mohamed, “Fake or Real? A Study of Arabic Satirical Fake News,” in Proceedings of the 3rd International Workshop on Rumours and Deception in Social Media (RDSM), Dec. 2020, pp. 70–80, [Online]. Available: https://aclanthology.org/2020.rdsm-1.7.

[90] M. A. Bsoul, A. Qusef, and S. Abu-Soud, “Building an Optimal Dataset for Arabic Fake News Detection,” Procedia Comput. Sci., vol. 201, pp. 665–672, 2022, doi: https://doi.org/10.1016/j.procs.2022.03.088.

[91] H. Himdi, G. Weir, F. Assiri, and H. Al-Barhamtoshy, “Arabic Fake News Detection Based on Textual Analysis,” Arab. J. Sci. Eng., 2022, doi: 10.1007/s13369-021-06449-y.

[92] S. Alqurashi, A. Alhindi, and E. A. Alanazi, “Large Arabic Twitter Dataset on COVID-19,” ArXiv, vol. abs/2004.0, 2020.

[93] M. Hasanain et al., “Overview of CheckThat! 2020 Arabic: Automatic identification and verification of claims in social media,” 2020.

[94] A. Chouigui, O. Ben Khiroun, and B. Elayeb, “ANT Corpus: An Arabic News Text Collection for Textual Classification,” in 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), 2017, pp. 135–142, doi: 10.1109/AICCSA.2017.22.

[95] Z. Guo, M. Schlichtkrull, and A. Vlachos, “A Survey on Automated Fact-Checking,” Trans. Assoc. Comput. Linguist., vol. 10, pp. 178–206, 2022, doi: 10.1162/tacl_a_00454.

[96] H. Mubarak and S. Hassan, “ArCorona: Analyzing Arabic Tweets in the Early Days of Coronavirus (COVID-19) Pandemic,” in Proceedings of the 12th International Workshop on Health Text Mining and Information Analysis, Apr. 2021, pp. 1–6, [Online]. Available: https://aclanthology.org/2021.louhi-1.1.

[97] Z. Sheikh Ali, W. Mansour, T. Elsayed, and A. Al - Ali, “AraFacts: The First Large Arabic Dataset of Naturally Occurring Claims,” in Proceedings of the Sixth Arabic Natural Language Processing Workshop, Apr. 2021, pp. 231 - 236, [Online]. Available: https://aclanthology.org/2021.wanlp-1.26.

[98] R. Baly, M. Mohtarami, J. Glass, L. Màrquez, A. Moschitti, and P. Nakov, “Integrating Stance Detection and Fact Checking in a Unified Corpus,” in Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers), Jun. 2018, pp. 21–27, doi: 10.18653/v1/N18-2004.

[99] T. Alhindi, A. Alabdulkarim, A. Alshehri, M. Abdul-Mageed, and P. Nakov, “AraStance: A Multi-Country and Multi-Domain Dataset of Arabic Stance Detection for Fact Checking,” in Proceedings of the Fourth Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda, Jun. 2021, pp. 57–65, doi: 10.18653/v1/2021.nlp4if-1.9.

[100] A. Khalil, M. Jarrah, M. Aldwairi, and M. Jaradat, “AFND: Arabic fake news dataset for the detection and classification of articles credibility,” Data Br., vol. 42, p. 108141, 2022, doi: https://doi.org/10.1016/j.dib.2022.108141.

[101] H.-Y. Lu, C. Fan, X. Song, and W. Fang, “A novel few-shot learning based multi-modality fusion model for COVID-19 rumor detection from online social media.,” PeerJ. Comput. Sci., vol. 7, p. e688, 2021, doi: 10.7717/peerj-cs.688.

[102] D. Wadden et al., “Fact or Fiction: Verifying Scientific Claims,” in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Nov. 2020, pp. 7534–7550, doi: 10.18653/v1/2020.emnlp-main.609.

[103] O. Sagi and L. Rokach, “Ensemble learning: A survey,” Wiley Interdiscip. Rev. Data Min. Knowl. Discov., vol. 8, no. 4, pp. 1–18, 2018, doi: 10.1002/widm.1249.