Arabic Misinformation Detection on Social Media: Methods, Challenges, and Future Directions

Esraa Hamdy
Department of Computer Science, Faculty of Graduate Studies for Statistical Researches, Cairo University, Egypt

Tarek Elghazaly
Department of Computer Science, Faculty of Graduate Studies for Statistical Researches, Cairo University, Egypt

Sarah Saad Eldin
Department of Computer Science, Faculty of Graduate Studies for Statistical Researches, Cairo University, Egypt

ABSTRACT

In the digital age, the widespread dissemination of misinformation and fake news poses significant threats to public opinion, social trust, and decision-making processes. This survey provides a comprehensive overview of recent advances in fake news detection, with a particular focus on techniques involving preprocessing, text representation, and classification. We further examine related tasks such as claim detection and veracity prediction, which form the backbone of automated fact-checking systems. The survey also explores some of the Arabic fake news detection datasets that have been published for various tasks. This work aims to offer valuable insights into current research trends, identify existing challenges, and highlight potential directions for future developments in the fight against misinformation.

Keywords

Misinformation detection – Arabic Fake News – Machine Learning – Automated Fact checking – Social Media

1. INTRODUCTION

The increasing growth of digital media has revolutionized the way information is produced, disseminated, and consumed. Social media platforms, in particular, have enabled rapid sharing of content on a global scale, granting individuals immediate access to news and opinions. However, this transformation has also created a fertile ground for the spread of fake news or false information presented as legitimate news. The widespread circulation of such content has posed significant risks to democratic institutions, public health, and societal trust, especially during sensitive periods such as elections and pandemics [1].

While most fake news detection research has been conducted on English-language, relatively limited attention has been given to low-resource languages such as Arabic. The Arabic language presents unique challenges due to its rich morphology, complex syntax, diglossia (the coexistence of Modern Standard Arabic and various dialects), and lack of large-scale annotated datasets [2]. Furthermore, the limited availability of pretrained language models tailored to Arabic content hinders the development of robust detection systems.

In response to the challenges posed by Arabic fake news, the research community has proposed various technological solutions to automatically detect and mitigate its impact. Broadly speaking, these solutions fall into two primary categories: fake news detection and automated fact-checking. The first line of research, fake news detection, typically treats the problem as a binary or multi-class classification task, where an entire news article or social media post is labeled as "fake" or "real" based on linguistic, semantic, and contextual features. These models often employ machine learning, deep learning, and transformer-based architectures to identify deceptive patterns in text. In contrast, automated fact-checking systems verify the truthfulness of claims by retrieving evidence from trusted sources and assessing their relationship [3].

Despite notable advancements in various domains, several key challenges remain, including the scarcity of high-quality annotated Arabic datasets and the fast-evolving nature of misinformation. This survey provides a comprehensive and unified analysis of the current landscape of fake news detection by bridging two major research directions: fake news detection and automated fact-checking approaches. While previous surveys have addressed these areas in isolation, this work uniquely integrates them to highlight their interdependencies and combined role in combating misinformation.

The key contributions of this survey are as follows:

- We give a clear overview of how misinformation is detected using both machine learning and deep learning.
- We explain how automated fact-checking helps in identifying false information.
- We provide comparative discussions of Arabic misinformation datasets.
- We identify open challenges in the field, such as data scarcity and model generalizability, and propose directions for future work.

Through this multifaceted contribution, our survey serves as a foundational reference for researchers and practitioners seeking to design more robust, secure, and context-aware systems for fake news detection.

2. BACKGROUND

This section presents the background and motivation behind the task of misinformation detection, along with recent developments in the field.

2.1 Definition and Characteristics of Fake News

Broadly, fake news refers to false or misleading information that is presented as legitimate news with the intent to deceive or manipulate the audience. Scholars have proposed various taxonomies for categorizing fake news [4]. A common framework divides it into several types, including:

- Disinformation refers to deliberately false information spread with the explicit aim of misleading or manipulating people, often to promote a biased agenda
- Misinformation refers to incorrect or misleading

information that spreads unintentionally, often due to honest mistakes or outdated knowledge, without the intention to deceive.

- Rumors involve unverified claims circulated among individuals, which may eventually be confirmed or disproven.
- Hoaxes are intentionally fabricated stories designed to deceive people into believing something that is not true.
- Fake news refers to false or misleading news. It mimics real news stories but is intentionally misleading or fabricated.
- satire and parody use irony and humor to entertain or criticize, but they can also mislead if their context is not understood, potentially spreading confusion if shared as factual information

In summary, fake news is not simply the opposite of truth—it is a dynamic, adaptive, and context-dependent phenomenon that exploits human psychology and technological infrastructure. Accurately defining and characterizing it is a crucial first step toward developing effective detection and mitigation strategies.

2.2 Impact of Fake News on Society

The proliferation of fake news has emerged as a significant societal threat, affecting various domains, including politics, public health, the economy, and social cohesion. Unlike traditional misinformation, which may be unintentional or benign, fake news is often deliberately crafted to mislead, manipulate, or provoke specific reactions. Its rapid dissemination—particularly through social media platforms—has amplified its reach and impact, enabling false information to spread faster and farther than verified news. One of the most critical areas affected is the political landscape. Fake news has been used for voter manipulation, such as the interference in the 2016 U.S. presidential election, which undermined democratic processes and eroded public trust in institutions.

In the realm of public health, fake news poses serious dangers. During the COVID-19 pandemic, the spread of health-related misinformation—ranging from conspiracy theories about the virus's origins to false claims about vaccines—contributed to confusion, fear, and vaccine hesitancy. These narratives not only hindered effective crisis response but also endangered lives by promoting harmful behaviors [4]. Economically, fake news has been linked to market manipulation, damage to corporate reputations, and consumer distrust. False reports about company earnings or product recalls can lead to fluctuations in stock prices, while misleading content in sectors like finance or real estate can influence investment decisions with potentially severe consequences.

On a broader societal level, the constant exposure to disinformation erodes social cohesion and trust. It creates an environment where distinguishing truth from falsehood becomes increasingly difficult, leading to skepticism toward credible sources and institutions. This information disorder can weaken collective decision-making and impede the formation of informed public opinion.

2.3 Fake News Detection vs. Fact Checking

Fake news detection and automated fact-checking are two closely related tasks aimed at combating misinformation. Although both approaches share the ultimate goal of determining the truthfulness of information, they differ significantly in methodological pipelines.

2.3.1 Fake news detection pipeline

The fake news detection pipeline, as illustrated in Figure 1, typically begins with a pre-processing phase, which includes general steps such as tokenization, removing stopwords and punctuation. For Arabic text, additional language-specific steps are applied, such as removing diacritics, normalizing similar letters (e.g., “أ،إ،آ” to “ا”), handling dialectal variations, and performing morphological analysis using tools like Farasa or CAMeL Tools. After preprocessing, text is represented using either traditional methods like Word2Vec and GloVe or contextual embeddings from transformer-based models such as MARBERT. These representations are then used as input for various classifiers, including traditional machine learning algorithms like SVM, deep learning models such as CNNs and LSTMs, or advanced transformer-based architectures [5]. Finally, the output can be a binary classification (e.g., fake or real) or a multiclass prediction that includes categories like satire, hoax, or true.

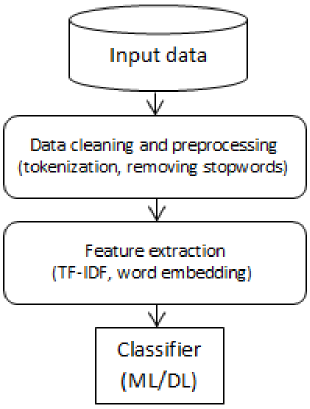

Fig 1: Fake news detection pipeline

2.3.2 Fact-checking pipeline

According to this study [3] fact-checking pipeline often composed of five steps, beginning with claim detection, where factual claims are automatically identified from raw sentences using techniques. Once claims are extracted, claim prioritization is applied to rank or filter the most check-worthy statements using scoring models or classifiers. In the next stage, evidence retrieval aims to find supporting or refuting evidence from trustworthy sources—such as knowledge graphs, fact-checking databases, or web searches—by transforming the claim into a query and using information retrieval techniques like BM25 or dense embeddings. Finally, in the veracity prediction phase, the system determines the truthfulness of the claim by analyzing the relationship between the claim and the retrieved evidence. Figure 2 illustrates the summarization of these steps.

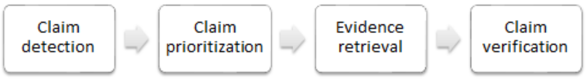

Fig 2: Fact-checking Pipeline.

3. SURVEY METHODOLOGY

The studies reviewed in this paper have been published in reputable journals and presented at international conferences by leading scientific publishers such as IEEE, Springer, Elsevier, Hindawi, Frontiers, Taylor & Francis, and MDPI. The relevant literature was identified using academic search engines, including Google Scholar, Scopus, and ScienceDirect.

The selected studies span the period from 2020 to 2025, reflecting recent advancements and trends in the field. This survey is structured into two main parts in the related work section: the first part reviews fake news detection approaches, while the second part focuses on claim verification and fact-checking techniques.

3.1 Fake News Detection

3.1.1 Machine learning approach

The authors of [6] proposed a hybrid model for Arabic fake news detection using NLP, ML, and Harris Hawks Optimizer (HHO) for feature selection. Using BoW and TF-IDF on a dataset of 1,862 tweets, their approach combining LR with TF-IDF and HHO improved feature selection and achieved a 5% performance gain over previous methods.

The authors of [7] proposed a two-phase method for detecting Arabic COVID-19 rumors. In the detection phase, a hybrid model combining Stacking Classifier (LR) and GA-SVM achieved 92.63% accuracy and strong precision, recall, and F1 scores on the Arcov-19 dataset. In the tracking phase, similarity measures like Cosine, Jaccard, and Chebyshev (with GloVe embeddings) were tested, with Chebyshev similarity performing best in ROUGE L scores.

Himdi et al. [8] develop a supervised machine learning model to detect Arabic fake news within a specific domain, namely the Hajj. They used Naïve Bayes (NB), Random Forest (RF), and SVM algorithms for credibility assessment. The model achieved over 75% accuracy, with linguistic features proving to be the most significant indicators.

The authors of [9] proposed a fusion method combining NER features with a novel feature selection algorithm, RLTTAO, which enhances TTAO using reinforcement and opposition-based learning. This approach improved accuracy by an average of 1.62% in 5 out of 7 datasets and outperformed other swarm optimization techniques.

The authors of [10] proposed a powerful Arabic fake news detection method combining advanced text features with a stacking classifier using bagging, boosting, and baseline models. Using ELMo embeddings and ensemble learning, their model achieved 99% accuracy, outperforming state-of-the-art methods.

3.1.2 *Deep learning approach*

The authors of [11] proposed a deep learning method combining CNN and LSTM to analyze the relationship between news headlines and articles using a dataset on the Syrian war and Middle Eastern politics. Their AFND-CNN-LSTM model achieved 70% accuracy, outperforming the LSTM-only model’s 68.2%.

The authors of [5] proposed JointBERT, integrating NER and relation fact classification with a unified BERT encoder for Arabic fake news detection. Evaluated on four datasets, it outperformed models like Qarib and AraBERT, achieving F1 scores of 85%, 66%, and 55% on Covid19Fakes, ANS, and Satirical datasets, respectively. The model improved average F1 by 10%, showing promise but highlighting the need for further research with long-term and self-supervised learning.

The authors of [1] develop transformer-based classifiers using eight Arabic models. The models were trained on two datasets: one collected from Twitter and the other translated from English. The models trained on the Arabic data outperformed those trained on translated data, with ARBERT and Arabic-BERT achieving accuracies of 98.8% and 98%, respectively.

The authors of [12] proposed a Transformer-based Multi-Task Learning model combined with the Nutcracker Optimization Algorithm for feature selection. Using fine-tuned AraBERT, the framework extracts rich contextual features from Arabic posts, achieving 87% accuracy in binary and 69% in multi-class classification, outperforming other methods and showing strong potential for misinformation detection.

The authors of [13] introduced a framework for detecting COVID-19 disinformation using multi-task learning, AraBERT, and a modified Fire Hawk Optimizer for feature selection. The two-phase approach improved performance, achieving 59% accuracy, and outperforming other methods in precision (53%), recall (71%), and F1 score (53%) across all datasets.

Albalawi et al. [14] propose a multimodal model for detecting rumors in Arabic. The model uses MARBERTv2 for textual feature extraction and an ensemble of VGG-19 and ResNet50 for visual feature extraction. The results show that unimodal text-based models outperform the multimodal models, highlighting the importance of textual features in Arabic rumor detection.

The authors of [15] proposed HPOHDL-FND, a hybrid deep learning model optimized with Hunter-Prey Optimization for Arabic fake news detection. Using extensive preprocessing and an LSTM-RNN classifier, the model achieved high accuracy of 96.57% on Covid19Fakes and 93.53% on satirical datasets.

The authors of [16] introduced AraCovTexFinder, a system for detecting Arabic COVID-19 texts using a fusion-based transformer model. It addresses challenges like limited data and dialect variation. With two new datasets (AraEC and AraCoV), the model achieved 98.89% accuracy, outperforming several transformer and deep learning models.

The authors of [2] developed a deep learning approach using transformer models on the ArabicFakeNews dataset. They found that shorter "description" texts gave the best detection results. AraBERTv2 achieved top performance with 97% accuracy, F1 score, and precision, and 96.58% recall, outperforming other deep learning and zero-shot models.

The authors of [17] proposed a hybrid model that combines Arabic BERT models (e.g., AraBERT, GigaBERT, MARBERT) with CNNs for fake news detection. Features are extracted using BERT, reduced via 1D or 2D-CNN, and classified using a neural network. The AraBERT + 2D-CNN model achieved the best results, with up to 71% accuracy and strong F1-scores, all in just two training epochs.

The authors of [18] proposed a deep ensemble framework using three BERT-based models to classify ten categories of Arabic COVID-19 content. Using the ArCOVID-19Vac dataset and a DAL technique with back translation and random insertion, they expanded the data and improved performance. The final model showed strong results across five augmented datasets.

Aljohani [19] addresses the challenge of an imbalanced dataset by using class weights, random under-sampling, SMOTE, and SMOTEENN techniques across several machine learning models such as XGBoost, Random Forest, CNN, BIGRU, BILSTM, CNN-LSTM, and CNN-BIGRU. The results demonstrate that SMOTEENN significantly improves model performance, particularly in terms of F1-score, precision, and recall.

Al-Zahrani and Al-Yahya [20] examine transformer-based ensemble models for Arabic fake news detection, utilizing five Arabic transformer models with various ensemble techniques.

Their findings show that ensemble models outperform baselines, achieving a 94% F1 score on the AMFND dataset.

The authors of [21] proposed a two-stage framework for Arabic fake news detection on Twitter, combining feature extraction and back-translation for data augmentation. Results showed that augmentation improved accuracy by 5–12%, with bi- and tri-grams outperforming traditional features. Medium-BERT further boosted performance, highlighting the value of data augmentation and deep learning.

The authors of [22] investigated ML, DL, and ensemble methods for Arabic fake news detection using FastText and transformer models. Their best-performing model, Bi-GRU-Bi-LSTM, achieved F1 scores of 0.98 and 0.99 on the AFND and ARABICFAKETWEETS datasets, proving highly effective in detecting fake news.

The authors of [23] introduced WaraBERT, a hybrid feature extraction method for fake news detection. Using AFND and AraNews datasets, WaraBERT-V1 achieved 93.83% accuracy with BiLSTM, while WaraBERT-V2 reached 81.25%. The study showed WaraBERT outperforms traditional methods like TF-IDF and AraBERT and highlighted the importance of keeping stopwords and Tanween marks for better accuracy. A summary of the previous studies is presented in Table 1.

Table 1 shows the previous studies of fake news detection

Ref.	Purpose	Approach	Dataset	Results
[6]	General false Arabic information detection	Various ML with feature selection techniques	Arabic tweets dataset (1862 instances)	5% improvement over previous methodson the same dataset.
[7]	COVID 19 Arabic Rumors detection	Stacking Classifier (LR) + GA-SVM	Arcov-19	Accuracy: 92.63%, Precision: 92.93%, Recall: 93.01%, F1: 92.83%
[8]	Arabic fake news detection	NB, RF, SVM with Arabic NLP features	Created Hajj news dataset	Accuracy: >75%
[9]	Arabic fake news detection	Proposes fusion of NER features with a novel feature selection algorithm (RLTTAO) with machine learning algorithms	7 datasets such as: OSACT, ArCOV19, and FKD	Avg. accuracy increase: 1.62% across 5/7 datasets
[10]	Arabic fake news detection	ELMO + Stacking (Bagging, Boosting, Baseline classifiers)	The Arabic Fake News Dataset (AFND)	Accuracy: 99%
[11]	Combines CNN and LSTM to Arabic fake news detection	Combines CNN and LSTM	They created Syrian war & Middle East claims dataset (422 claims, 3,042 articles)	Accuracy: 70% vs. 68.2%
[5]	Arabic fake news detection	Proposes JointBERT integrating NER and RFC with a unified encoder	Covid19Fakes, ANS, Satirical datasets	F1: 85%, 66%, 55%; Avg. 10% F1 improvement
[1]	Arabic fake news detection	Compares performance of 8 Arabic transformer-based	Twitter dataset + translated dataset	Accuracy: ARBERT (98.8%), Arabic-BERT (98%)
[12]	Arabic misinformation detection	Introduces Transformer-based MTL with NOA for feature selection (AraBERT + MTL + NOA)	utilizing diverse datasets(textual data and Twitter dataset)	Binary Accuracy: 87%, Multi-class: 69%
[13]	COVID-19 disinformation detection	MTL, AraBERT, and Fire Hawk Optimizer	COVID-19 fake news (multi-dataset)	Accuracy: 59%, Precision: 53%, Recall: 71%, F1: 53%
[14]	Multimodal rumor detection	For text (MARBERTv2) and image (VGG-19 + ResNet50)	Arabic multimodal rumor datasets	Text-only outperformed multimodal model
[15]	Arabic fake news detection	Proposes HPOHDL-FND using LSTM-RNN and Hunter-Prey Optimization	Covid19Fakes & Satirical datasets	Accuracy: 96.57%, 93.53%
[16]	COVID-19 misinformation detection	Introduces AraCovTexFinder using fusion-based transformer	AraEC, AraCoV	Accuracy: 98.89%
[2]	Arabic Fake News detection	AraBERTv2 and other transformers	ArabicFakeNews dataset (2,000 fake news items)	Accuracy 97%, F1 97%, Precision 97%, Recall 96.58%
[17]	Arabic Fake News detection	Hybrid approach combining Arabic BERT models with CNNs (AraBERT, GigaBERT, MARBERT + 1D-CNN/2D-CNN + ANN)	ANS dataset	AraBERT + 2D-CNN best; F1 scores up to 0.8009, 71% accuracy in 2 epochs; reduced training time
[18]	COVID-19 disinformation detection	Ensemble of 3 BERT models + Data Augmentation and Labeling (DAL), back translation, random insertion	ArCOVID-19Vac dataset	Impressive results across 5 augmented datasets (no exact metrics given)
[19]	Arabic fake news detection	XGBoost, Random Forest, CNN, BiGRU, BiLSTM, CNN-LSTM, CNN-BiGRU	Various Arabic fake news datasets	SMOTEENN improved F1, precision, recall significantly
[20]	Arabic fake news detection	Various Arabic transformers + ensemble techniques	AMFND dataset	Ensemble outperformed baselines; F1 score 94%

[21]	Arabic fake news detection	N-grams (bi, tri), Medium-BERT, data augmentation using back-translation	Arabic tweets dataset	Data augmentation improved accuracy by 5–12%; Medium-BERT enhanced performance
[22]	Arabic fake news detection	Bi-GRU-Bi-LSTM among others	AFND and ARABICFAKETWEETS datasets	Best model F1: 0.98 (AFND), 0.99 (ARABICFAKETWEETS); effective fake news detection
[23]	Arabic fake news detection	WaraBERT-V1 (BiLSTM), WaraBERT-V2	AFND and AraNews datasets	WaraBERT-V1 accuracy 93.83% (AFND), V2 81.25% (AraNews)

3.2 Fact-Checking

3.2.1 Claim detection and prioritization

The authors of [24] participated in CLEF CheckThat! 2020 Arabic task for detecting check-worthy tweets. They compared traditional classifiers with hand-crafted features to a multilingual BERT (mBERT) model, finding that mBERT outperformed traditional methods, ranking 6th overall and 3rd among eight teams.

The authors of [25] participated in CLEF-2020 CheckThat! Lab for check-worthiness detection in English and Arabic (Tasks 1 and 5). They combined fine-tuned BERT and AraBERT with logistic regression, using features like POS tags, controversial topics, word lists, and embeddings. Their models ranked 3rd in Arabic Task 1, 5th in English Task 1, and 3rd in Task 5.

The authors of [26] participated in CLEF2021 CheckThat! Lab Task 1, targeting check-worthy claim detection in five languages, including Arabic. They used deep transformer models enhanced with context-sensitive lexical data augmentation to expand training data. This approach notably boosted performance, with their Arabic system achieving the highest mean average precision in the lab.

The authors of [27] participated in CheckThat! 2022 for Arabic check-worthy tweet detection, fine-tuning several pre-trained Arabic transformers, including AraBERT, ARBERT, MARBERT, Arabic ALBERT, and BERT base Arabic. AraBERT achieved the highest F1 score (0.462) on Subtask 1A, while ARBERT led Subtask 1C with an F1 of 0.557.

The authors of [28] explored GPT-3 for automatic claim detection in CheckThat! 2022. GPT-3 achieved the best accuracy (0.761) in subtask 1B and ranked third in subtask 1A (F1-positive: 0.626) without preprocessing or augmentation. However, its performance was weaker for Arabic and Bulgarian and notably lower in subtasks 1C and 1D due to challenges in class label identification.

The authors of [29] used transformer models, including XLM-RoBERTa, to detect check-worthy claims in multiple languages at CheckThat! 2023. They ranked 3rd in Arabic and 5th in English for Subtask 1A, and 3rd in Arabic and 6th in English and Spanish for Subtask 1B.

The authors of [30] combined fine-tuned language and vision transformers with BiLSTM and multi-sample dropout for multimodal check-worthiness detection. Their approach ranked 1st in Arabic multimodal and 3rd in Spanish subtasks at CheckThat! 2023.

The authors of [31] participated in CLEF2024 CheckThat! Lab Task 1 using generative transformers with few-shot reasoning, fine-tuning, data augmentation, and cross-lingual transfer. They ranked 1st in Arabic (F1=0.569), 3rd in Dutch (F1=0.718), and 9th in English (F1=0.753).

The authors of [32] participated in CLEF2024 CheckThat! Lab Task 1, fine-tuning mono- and multilingual models with data balancing and cross-lingual transfer learning (without instance transfer). Their method ranked 2nd in Arabic (F1=0.557) and English (F1=0.796). They found that handling class imbalance was more important than annotation quality.

3.2.2 Claim veracity

The authors of [33] participated in CLEF2020 CheckThat! lab on Arabic tasks, including check-worthy tweet detection, evidence retrieval, and claim verification. Their system combined sentiment, linguistic, and named entity features with cosine similarity. Despite sentiment features initially lowering performance, they improved final classification, achieving an F1-score of 0.55. They also explored unsupervised methods to reduce annotation needs and speed fact-checking.

The authors of [34] created AuRED, the first Arabic Authority-Rumor-Evidence Dataset with 160 rumors and 692 authoritative timelines (≈34,000 annotated tweets). Experiments showed strong evidence retrieval performance, including cross-lingual zero-shot, but weak rumor verification results. The study highlights the value of stance detection and transfer learning from other fact-checking datasets.

The authors of [35] introduced Ta'keed, an explainable Arabic fact-checking system combining information retrieval and an LLM-based verifier. Evaluated on the ArFactEx dataset, it achieved an F1 score of 0.72, with explanations scoring 0.76 in semantic similarity. Using the top seven snippets improved accuracy further, reaching an F1 of 0.77. Table 2 summarizes the related works on Arabic fact-checking approaches.

4. DATASETS

This section presents some of the Arabic fake news detection datasets that have been published for various tasks. The authors of [36] presented a public Arabic corpus for credibility classification containing about 1570 credible and 1138 non-credible tweets. The authors of [37] built a novel Arab corpus for fake news detection that contains 4,079 news items (3,286 not-rumor and 793 rumor). The authors of [38] created the'AraCOVID19-MFH' dataset for Arabic COVID-19 misinformation and hate speech detection. The KUNUZ collection consists of 10,828 Arabic tweets; 459 are fake. The authors of [39] created ArCOV19-Rumors dataset for health claims and other topics such as social, political, sports, entertainment, and religious. ArCOV19-Rumor is a collection of 1,753 false, 1,831 true, and 5,830 other tweets. The authors of [40] built a new Arabic stance detection of 4,063 claim articles of various domains (e.g., politics, sports, health). The authors of [41] created a dataset of COVID-19 misinformation. Their corpus consists of 8786 tweets annotated as misinformation or not, which will be freely available for the research community.

5. CHALLENGES AND FUTURE WORK

Despite the promising results, several challenges remain in the domain of misinformation detection. First, the scarcity of large-scale, multi-class labeling annotated datasets in low-resource languages, such as Arabic, significantly limits the generalizability of current models. Moreover, there is limited diversity in news domains, as most existing datasets mainly focus on COVID-19-related content. Additionally, most existing datasets also suffer from class imbalance, with the real news class significantly outweighing the fake news class, which can bias model predictions. Furthermore, there is a lack of Arabic multimodal datasets to support the development of multimodal misinformation detection models.

For future work, we plan to explore more advanced and recent models to enhance misinformation detection performance. Incorporating multi-modal information (e.g., images, videos) alongside textual data could also enhance detection performance. Additionally, integrating metadata such as publisher names, URLs, headlines, article bodies, and user comments can provide deeper context for detecting fake news more accurately.

6. CONCLUSION

In conclusion, the detection of false information in the Arabic language presents unique challenges due to its rich morphology, the presence of numerous dialects, and the limited availability of labeled datasets and pre-trained models. While substantial advancements have been achieved in other languages, Arabic remains underrepresented in both research and resources. The emergence of Arabic fact-checking platforms such as Fatabyyano and Misbar, has been instrumental in providing structured annotations and raising public awareness. Nevertheless, more efforts are needed to develop robust Arabic-specific tools, expand annotated corpora, and adopt cross-lingual techniques to enhance misinformation detection in this low-resource language. Addressing these challenges is essential for building effective, culturally aware systems capable of combating Arabic misinformation. Future research could focus on developing hybrid approaches that integrate content and context signals, leveraging advances in large language models, and creating benchmark datasets for Arabic language. Such directions may contribute to building more robust and generalizable Arabic misinformation detection systems.

Table 2 shows the previous studies of fact-checking approach

Ref.	Purpose	Approach	Dataset	Results
[24]	for ranking Arabic tweets by check-worthiness	Multilingual BERT (mBERT) model & Logistic Regression, Support Vector Machine (SVM) and Random Forest	CLEF CheckThat! Lab 2020 Arabic task	Ranked 6th overall; mBERT outperformed traditional methods
[25]	Check-worthiness detection	BERT + AraBERT + logistic regression + linguistic features	CLEF 2020 CheckThat! Lab (English and Arabic)	3rd place Arabic Task 1
[26]	identifying and ranking check-worthy claims in social media posts	Deep neural transformers + lexical data augmentation	CLEF 2021 CheckThat! Lab, multi-language	Highest MAP for Arabic; augmentation improved performance notably
[27]	Harmful tweet detection and Check-worthiness of tweets	AraBERT, ARBERT, MARBERT, Arabic ALBERT, BERT base Arabic	Subtask 1A: Check-worthiness of tweets& Subtask 1C: Harmful tweet detection	AraBERT F1=0.462 (Subtask 1A); ARBERT F1=0.557 (Subtask 1C)
[28]	GPT-3 for claim detection	GPT-3 zero/few-shot classification	CheckThat! 2022 dataset	GPT-3 outperformed BERT
[29]	Unimodal and multimodal check-worthiness classification using transformers and vision models	XLM-RoBERTa-large + BERT + ResNet50 + feed-forward networks	CLEF 2023 CheckThat! Lab multimodal/multigenre tweets	3rd Arabic & 5th English in Subtask 1A; For Subtask 1B, they ranked 3rd in Arabic and 6th in English and Spanish.
[30]	Fusion of language and vision transformers with BiLSTM + multi-sample dropout for check-worthiness	transformer models including XLM-RoBERTa3, AraBERT4, BERTweet5, Spanish BERT6 and ConvNEXT7 model	CLEF 2023 CheckThat! Lab multimodal/multigenre tweets	1st Arabic multimodal; 3rd Spanish;
[31]	Transformer models + data augmentation + few-shot reasoning for check-worthiness	GPT-3.5, GPT-4, RoBERTa-Large + cross-lingual transfer	CLEF 2024 CheckThat! Lab (English, Dutch, Arabic)	1st Arabic (F1=0.569), 3rd Dutch (F1=0.718), 9th English (F1=0.753)
[32]	Automated approach for identifying check-worthy claims	Mono/multilingual PLMs (CAMeLBERT MSA mDeBERTa V3)	CLEF 2024 CheckThat! Lab (English, Dutch, Arabic)	0.557 F1 score for Arabic 0.590 F1 score for Dutch 0.796 F1 score for English
[33]	identifying check-worthy tweets, retrieving evidence from the web, and verifying claims	Sentiment, linguistic features, named entities, cosine similarity for claim-evidence alignment	CLEF2020 CheckThat! Lab (Arabic tasks)	F1=0.55 for classification task
[34]	Rumor verification using evidence from authoritative Twitter accounts; construct and release a dataset for Arabic low-resourced language.	Evidence Retrieval Models(BM25, mContriever, KGAT, MLA, TML, STAuRED), Rumor Verification Models(MLA, KGAT)	Authority-Rumor-Evidence Dataset (AuRED)	Strong evidence retrieval, poor rumor verification

[35]	Ta'keed: Explainable Arabic fact-checking system with justification generation	Info retrieval + LLM-based verification + explanation generation;	ArFactEx dataset	F1=0.72; for classification task cosine similarity score =0.76; best F1=0.77 using top 7 snippets

7. REFERENCES

[1] A. B. Nassif, A. Elnagar, O. Elgendy, and Y. Afadar, "Arabic fake news detection based on deep contextualized embedding models," Neural Computing and Applications, vol. 34, no. 18, pp. 16019-16032, 2022.

[2] I. kh. Alnabrisi and M. kh. Saad, "Detect Arabic fake news through deep learning models and Transformers," Expert Systems with Application, vol. 251, p. 123997, 2024.

[3] R. Panchendrarajan and A. Zubiaga, "Claim detection for automated fact-checking: A survey on monolingual, multilingual and cross-lingual research," Natural Language Processing Journal, vol. 7, p. 100066, 2024.

[4] T. Alotaibi and H. Al-Dossari, "A Review of Fake News Detection Techniques for Arabic Language," Int. J. Adv. Comput. Sci. Appl., vol. 15, no. 1, pp. 392–407, 2024.

[5] W. Shishah, "JointBert for Detecting Arabic Fake News," IEEE Access, vol. 10, no. July, pp. 71951–71960, 2022.

[6] 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, vol. 13, no. 4, p. 556, 2021.

[7] S. N. Qasem, M. Al-Sarem, and F. Saeed, "An ensemble learning based approach for detecting and tracking COVID19 rumors," Comput. Mater. Contin., vol. 70, no. 1, pp. 1721–1747, 2021.

[8] H. Himdi, G. Weir, F. Assiri, and H. Al-Barhamtoshy, "Arabic fake news detection based on textual analysis," Arab. J. Sci. Eng., vol. 47, no. 8, pp. 10453–10469, 2022.

[9] A. Dahou et al., "Linguistic feature fusion for Arabic fake news detection and named entity recognition using reinforcement learning and swarm optimization," Neurocomputing, vol. 598, no. March, p. 128078, 2024.

[10] T. Aljrees, "Improving Prediction of Arabic Fake News Using ELMO's Features-Based Tri-Ensemble Model and LIME XAI," IEEE Access, vol. 12, no. February, pp. 63066–63076, 2024.

[11] H. Najadat, M. Tawalbeh, and R. Awawdeh, "Fake news detection for Arabic headlines-articles news data using deep learning," Int. J. Electr. Comput. Eng., vol. 12, no. 4, pp. 3951–3959, 2022.

[12] A. Dahou et al., "Optimizing fake news detection for Arabic context: A multitask learning approach with transformers and an enhanced Nutcracker Optimization Algorithm," Knowledge-Based Systems, vol. 280, p. 111023, 2023.

[13] 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, pp. 1–15, 2023, doi: 10.3390/math11020258.

[14] R. M. Albalawi, A. T. Jamal, A. O. Khadidos, and A. M. Alhothali, "Multimodal Arabic Rumors Detection," IEEE Access, vol. 11, no. December 2022, pp. 9716–9730, 2023, doi: 10.1109/ACCESS.2023.3240373.

[15] H. J. Alshahrani et al., "Hunter Prey Optimization with Hybrid Deep Learning for Fake News Detection on Arabic Corpus," Computers, Materials & Continua, vol. 75, no. 2, 2023.

[16] M. R. Hossain, M. M. Hoque, N. Siddique, and M. A. A. Dewan, "AraCovTexFinder: Leveraging the transformer-based language model for Arabic COVID-19 text identification," Engineering Applications of Artificial Intelligence, vol. 133, p. 107987, 2024.

[17] N. A. Othman, D. S. Elzanfaly, and M. M. M. Elhawary, "Arabic Fake News Detection Using Deep Learning," IEEE Access, 2024.

[18] A. Y. Muaad, H. J. Davanagere, J. Hussain, and M. A. Al-antari, "Deep ensemble transfer learning framework for COVID-19 Arabic text identification via deep active learning and text data augmentation," Multimedia Tools and Applications, vol. 83, no. 33, pp. 79337-79375, 2024.

[19] E. Aljohani, "Enhancing Arabic Fake News Detection: Evaluating Data Balancing Techniques Across Multiple Machine Learning Models," Engineering, Technology & Applied Science Research, vol. 14, no. 4, pp. 15947-15956, 2024.

[20] L. Al-Zahrani and M. Al-Yahya, "Pre-trained language model ensemble for Arabic fake news detection," Mathematics, vol. 12, no. 18, pp. 1-17, 2024.

[21] E. A. Mohamed, W. N. Ismail, O. A. S. Ibrahim, and E. M. Younis, "A two-stage framework for arabic social media text misinformation detection combining data augmentation and arabert," Social Network Analysis and Mining, vol. 14, no. 1, p. 53, 2024.

[22] M. E. Almandouh, M. F. Alrahmawy, M. Eisa, M. Elhoseny, and A. Tolba, "Ensemble based high performance deep learning models for fake news detection," Scientific Reports, vol. 14, no. 1, p. 26591, 2024.

[23] H. M. Turki et al., "Arabic fake news detection using hybrid contextual features," International Journal of Electrical & Computer Engineering (2088-8708), vol. 15, no. 1, 2025.

[24] M. Hasanain and T. Elsayed, "bigIR at CheckThat! 2020: Multilingual BERT for ranking Arabic tweets by check-worthiness," 2020.

[25] Y. S. Kartal and M. Kutlu, "TOBB ETU at CheckThat! 2020: Prioritizing English and Arabic Claims Based on Check-Worthiness," in CLEF (Working Notes), 2020.

[26] E. Williams, P. Rodrigues, and S. Tran, "Accenture at CheckThat! 2021: interesting claim identification and ranking with contextually sensitive lexical training data augmentation," arXiv preprint arXiv:2107.05684, 2021.

[27] B. Taboubi, M. A. B. Nessir, and H. Haddad, "iCompass at CheckThat!-2022: ARBERT and AraBERT for Arabic

Checkworthy Tweet Identification," in CLEF (Working Notes), 2022, pp. 702-709.

[28] S. Agresti, S. A. Hashemian, and M. J. Carman, "PoliMi-FlatEarthers at CheckThat!-2022: GPT-3 applied to claim detection," CLEF (Working Notes), vol. 2022, 2022.

[29] P. Tarannum, M. A. Hasan, F. Alam, and S. R. H. Noori, "Z-Index at CheckThat!-2023: Unimodal and Multimodal Check-worthiness Classification," in CLEF (Working Notes), 2023, pp. 482-493.

[30] A. Aziz, M. A. Hossain, and A. N. Chy, "CSECU-DSG at CheckThat!-2023: Transformer-based Fusion Approach for Multimodal and Multigenre Check-Worthiness," in CLEF (Working Notes), 2023, pp. 279-288.

[31] P. R. Aarnes, V. Setty, and P. Galuščáková, "Iai group at checkthat! 2024: Transformer models and data augmentation for checkworthy claim detection," arXiv preprint arXiv:2408.01118, 2024.

[32] M. Sawiński, K. Węcel, and E. Księżniak, "OpenFact at CheckThat! 2024: Cross-Lingual Transfer Learning for Check-Worthiness Detection," in CEUR Workshop Proceedings, 2024, vol. 3740.

[33] I. Touahri and A. Mazroui, "EvolutionTeam at CLEF2020-CheckThat! Lab: Integration of linguistic and sentimental features in a fake news detection approach," in CLEF (working notes), 2020.

[34] F. Haouari, T. Elsayed, and R. Suwaileh, "AuRED: enabling Arabic rumor verification using evidence from authorities over Twitter," in Proceedings of The Second Arabic Natural Language Processing Conference, 2024, pp. 27-41.

[35] S. Althabiti, M. A. Alsalka, and E. Atwell, "Ta'keed: The First Generative Fact-Checking System for Arabic Claims," arXiv preprint arXiv:2401.14067, 2024.

[36] A. Al Zaatari et al., "Arabic corpora for credibility analysis," in Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), 2016, pp. 4396-4401.

[37] F. A. Alqahtani and M. Sanderson, "Generating a lexicon for the Hijazi dialect in Arabic," in International Conference on Arabic Language Processing, 2019: Springer, pp. 3-17.

[38] M. S. H. Ameur and H. Aliane, "AraCOVID19-MFH: Arabic COVID-19 multi-label fake news & hate speech detection dataset," Procedia Computer Science, vol. 189, pp. 232-241, 2021.

[39] F. Haouari, M. Hasanain, R. Suwaileh, and T. Elsayed, "ArCOV19-rumors: Arabic COVID-19 twitter dataset for misinformation detection," arXiv preprint arXiv:2010.08768, 2020.

[40] 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," arXiv preprint arXiv:2104.13559, 2021.

[41] S. Alqurashi, B. Hamoui, A. Alashaikh, A. Alhindi, and E. Alanazi, "Eating garlic prevents COVID-19 infection: Detecting misinformation on the Arabic content of Twitter," arXiv preprint arXiv:2101.05626, 2021.