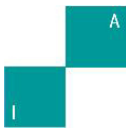
Inteligencia Artificial, 26(72) (2023), 178-201
doi: 10.4114/intartif.vol26iss72pp178-201



INTELIGENCIA ARTIFICIAL

http://journal.iberamia.org/

Fake News Detection in Low Resource Languages using SetFit Framework

Amin Abdedaiem [1,A], Abdelhalim Hafedh Dahou [2,C], Mohamed Amine Cheragui [1,B]

[1] Department of Mathematics and Computer Science, Ahmed Draia University, Adrar 01000, Algeria.
[A] aminabdedaiem@gmail.com , [B] m_cheragui@univ-adrar.edu.dz

[2] GESIS - Institute for Social Sciences, 50667 Cologne, Germany
[C] Abdelhalim.Dahou@gesis.org

Abstract Social media has become an integral part of people's lives, resulting in a constant flow of information. However, a concerning trend has emerged with the rapid spread of fake news, attributed to the lack of verification mechanisms. Fake news has far-reaching consequences, influencing public opinion, disrupting democracy, fueling social tensions, and impacting various domains such as health, environment, and the economy. In order to identify fake news with data sparsity especially with low resources languages such as Arabic and its dialects, we propose a few-shot learning fake news detection model based on sentence transformer fine tuning, utilizing no crafted prompts and language model with few parameters. The experimental results prove that the proposed method can achieve higher performances with fewer news samples. This approach provided 71% F1 score on the Algerian dialect fake news dataset and 70% F1 score on the Modern Standard Arabic (MSA) version of the same dataset, which prove that the approach can work on the standard Arabic and its dialects. Therefore, the proposed model can identify fake news in several domains concerning the Algerian community such as politics, COVID-19, tourism, ecommerce, sport, accidents and cars prices.

Keywords: Fake News, Algerian Dialect, Social Media, Deep Learning, Few-Shot Learning, Normalization, MSA, Arabic, BERT

1 Introduction

In the modern digital age, the internet has become pervasive, with 5.16 billion users worldwide, comprising 64.4 percent of the global population[1]. Social media platforms like Facebook, Snapchat, and Twitter have emerged as vital information hubs, witnessing enormous user engagement. Every minute, these platforms see a remarkable volume of comments, photos, and posts. Consequently, social media has become a preferred source of information for many, offering speed and cost-effectiveness compared to traditional news platforms. This shift highlights the increasing impact of social media on information spreading and consumption. The transition to social media as a primary source of information has a downside: the lack of control and verification makes it fertile ground for the rapid spread of unverified or false information [63]. Catchy titles alone can lead to thousands of instant shares without verifying sources or information. This has caused two significant issues: a surge in data volume and a shortage of tools to discern credible information from untrue, giving rise to fake news [9], false news [59], satire news [19], disinformation [39], misinformation [40], and rumors [24]. False information serves various purposes, like increasing website

[1] https://www.statista.com/statistics/617136/digital-population-worldwide/

ISSN: 1137-3601 (print), 1988-3064 (on-line)

or social media traffic and influencing public opinion on political decisions and financial transactions. As a consequence, the detection of fake news is undeniably a challenge in the fight against misinformation and false information and has become an area of interdisciplinary research due the complexity of the task, since rumors can be disseminated in a variety of forms, including texts, images, videos, and memes, and can be intentionally misleading in order to appear believable.

While there has been extensive research on identifying fake news in English text, there is a notable lack of studies focused on Arabic language social media platforms. This scarcity can be attributed to the considerable complexities within Arabic natural language processing, especially encompassing numerous dialects which are much distinct from both each other and the standard form (MSA).

The Arabic dialect referred to as ammiyya (common language) or darija (current language), this linguistic variant is a form of the Arabic language [18]. It is commonly employed in everyday communication across Arab countries [20]. Dialects, as defined by [33], represent the prevalent language used in daily activities and are typically spoken, though occasionally in written form. These dialects not only differ between various Arab territories but also exhibit variations from one region to another within the same territory. Among the 22 Arab countries, each possesses its unique dialect, which may share the same script but feature distinct pronunciations. As stated by [6], there are nine (09) distinct dialectal categories within the Arab world: Egyptian, Gulf (encompassing Bahrain, Kuwait, Qatar, Oman, and others), Iraqi, Levantine (covering Jordan, Lebanon, Palestine, and Syria), Maghrebian (encompassing Algeria, Tunisia, Morocco, and Libya), Yemeni, Somali, Sudanese, and Mauritanian. Within each of these categories, numerous varieties exist, depending on specific cities and regions.

In the Maghrebian dialects, Algerian dialect has not been sufficiently studied especially in the context of fake news, knowing that it is widely used in social networks. In Algeria, the linguistic landscape comprises two distinct forms of communication: MSA and the Algerian Dialect (ALG-DIA). MSA serves as the official and standardized version, used in formal contexts like media, newspapers, education, and official gatherings, closely resembling Classical Arabic. On the other hand, ALG-DIA represents the everyday spoken language, prevalent in daily life, advertising, music, and social media platforms like Facebook and Twitter among family and friends [34]. Despite its widespread use, ALG-DIA is considered a loosely standardized low-variety language, spoken by approximately 70-80% of the Algerian population, while the remaining 20-30% communicate in Berber [50]. This Algerian dialect exhibits regional variations, including:

- **Algiers dialect:** Common in Algiers and its vicinity.
- **Oran Dialect:** Spoken in western Algeria, from the Algerian-Moroccan border to Tenes (city of Chelf).
- **Rural dialect:** Used in eastern Algeria, encompassing areas like Constantine, Annaba, Setif, and regions near the Algerian-Tunisian border.
- **Saharan dialect:** Predominant among the population residing in southern Algeria.

The Algerian dialect has evolved from various languages, including French, Turkish, Spanish, Italian, and Portuguese, diverging significantly from MSA in terms of phonetics, phonology, morphology, and syntax. It lacks a standardized writing system, allowing flexible spellings understood by native speakers. Additionally, the dialect features linguistic phenomena as shown in table 1, like code-switching between Arabic and French within sentences and the Arabizi form by using Latin characters, numbers, and special characters instead of Arabic ones in its written style, reflecting its dynamic and evolving nature.

Table 1: Algerian dialect's writing style variations.

Arabizi	ycombati la misere li las9at fina welat kiste
Arabic transliteration (Dz)	يكومباطي لا ميزار لي لسقت فينا ولات كيست
Code-switched transliteration	kyste ولات فينا لسقت لي la misere يكومباطي
English translation	He fights the misery that stuck to us and which has become a cyst

The goal of this study is to identify both fake and real news in texts composed in the Algerian dialect. Given the absence of available datasets for detecting fake news in the Algerian language, we gathered our own data by performing web scraping on various social platforms. Subsequently, we formulated four research inquiries:

- **RQ1.** How does the variation in the number of training samples per class (in the context of Few-shot learning) impact the performance of different pre-trained models, including MARBERTv2, DziriBERT, AraBART, AraBERTv2, and DarijaBERT ? This study aims to compare their performance under varying training sample conditions.
- **RQ2.** When utilizing a pre-trained translation model from the Algerian dialect to MSA, how do the same models mentioned in the previous question perform on a normalized dataset ? This study aims to investigate the impact of normalization from Algerian dialect to MSA.
- **RQ3.** Can models trained on MSA, such as ARBERT and CAMeLBERT MSA, achieve satisfactory results when applied to the Algerian dialect ? It appears that the Algerian dialect has also developed from MSA. We aim to determine the extent of their linguistic commonalities.
- **RQ4.** In the fourth research, we assess and compare the performance of two architectural approaches: standard fine-tuned transformer model versus few-shot learning model.

The paper is structured as follows: In Section 2, we give an overview of different approaches that have been used in fake news detection. In Section 3, we describe the related work. We outline in Section 4 the methodology we used. In Section 5, we give the details of the experimental results with a discussion of important points noted during the different tests and conclude this paper with summary and a discussion of future work in Section 6.

2 Fake News Detection Methods

As mentioned previously, the spread of fake news has become a major concern due to the rise of social media and the ease of creating and sharing content. In recent years, there has been growing interest in developing methods to detect fake news and limit its spread. Fake news detection methods can include a variety of techniques that focus on several features. The aim of these methods is to help individuals and organizations identify and avoid false or misleading information, and ultimately, to help combat the spread of misinformation. In this section, we discuss various methods built for fake news detection by focusing on the common uses in literature such as content and social context of the news as shown in figure 1.

2.1 Content Based

The first kind of feature extraction phase for the fake news detection process is named content based. This method involves creating a structured representation of the news content, which can include the text of the article, images, or both. From this content, various types of features can be extracted, including linguistic features that focus on the language used, visual features that relate to the images or videos, and knowledge-based features that draw on external sources to assess the credibility of the content.
To detect fake news based on the content, it is crucial to analyze the linguistic features of the content. This involves examining various aspects, including the writing style, sentiment, and structure of the content [26]. In social media, fake news often uses biased and inflammatory language to mislead readers and achieve financial or political gains. To improve the accuracy of the fake news detection method, we can focus on the lexical features of the content such as the length of words, unique words, total word count, and frequency of longer words, as well as the syntactic features including sentence structure, punctuation, and parts-of-speech. By incorporating these linguistic features, we can enhance the predictive ability of the final models.
In order to mislead readers, social media users attached visual materials as evidence to increase the credibility of the news [25]. Verifying the visual material can help to address the problem based on some

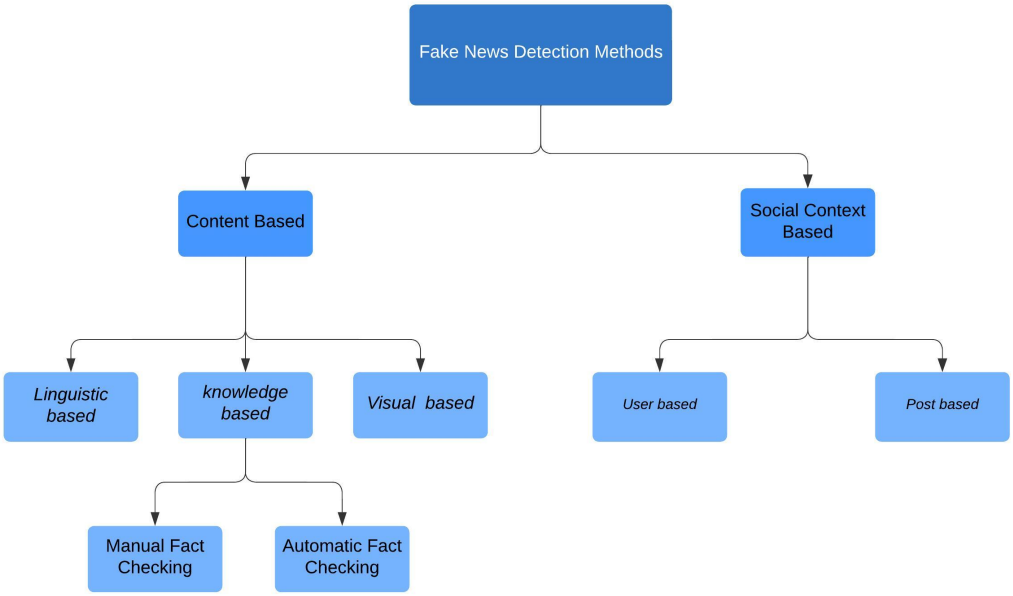


Figure 1: The most common fake news detection methods.

visual features such as extracted concepts, image captioning and clarity score. Both linguistic and visual features can suffer from limitations where the content is short in terms of text or content that does not contain any visual materials. The knowledge feature or approach uses fact checking techniques in order to compare the news content with several predefined external sources to determine the trustworthiness of the input news. The fact checking technique can be done by an expert in order to make a final decision about the news by manually checking external sources like *snopes* and *Poltifact* . Also can be done by a crowd sourced technique which helps to check the accuracy of the news but this one is less credible compared to the human expert [62].

2.2 Social Context Based

Based on [10] [37], the quality of news on social media is much lower than that of traditional news organizations. Large volumes of fake news are produced online for a variety of purposes, such as financial and political gain, and more people now are seeking and consuming information from social media rather than traditional news organizations. For that, social media acts as a source of auxiliary information for inferring the truthfulness of news pieces that takes into consideration the publisher, news piece and the users that spread the news and interact with. Most works in literature focus on extracting the features from users and posts to help fake news detection systems.
To detect fake news, it is also important to examine user engagement with news content on social media. Users on social media have varying levels of credibility, which can provide insights into their likelihood of sharing fake news. Credibility is determined by the degree to which a user is deemed trustworthy. The credibility criteria can vary depending on the specific context and methods used, but they generally include: (i) user history which involves examining their past behavior, such as the types of content they have shared or interacted with; (ii) the consistency of a user's behavior and posts over time is also taken into account. Users who consistently share accurate and reliable information are seen as more credible compared to those who frequently change their stance or spread conflicting information; (iii) verified users are generally considered more credible because their identity has been verified by the platform; (iv) users who share information from reputable sources or news organizations are often considered more credible. Users who have low credibility, including malicious accounts or individuals who are susceptible to misinformation, are more likely to spread fake news. Additionally, analyzing the interaction and opinions of users with a particular post on social media can serve as a useful feature to assess the credibility of the

news being shared.

3 Related works

In this Section, we are trying to find out the level reached with fake news detection in dialectical Arabic, whether it's the whole Arabic language (MSA) or dialectical Arabic related to a specific region or the Algerian dialect, we are also focusing on any existing datasets or systems that have been developed for the sake of fake news detection, we are trying to spot the problems researchers faced in this topic and the techniques used to encounter the spread of misinformative news. we have selected papers that searched in the context of fake news detection in Arabic in the last 7 years which covered all the possible approaches and methods from papers that used machine learning to ones that used transformers or hybrid techniques.the diversity of the papers selected is described in figure 2.

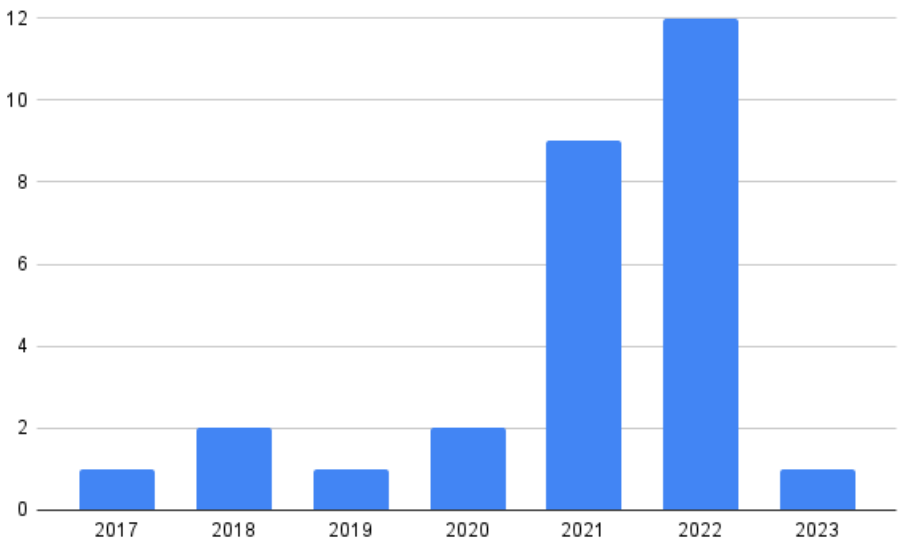


Figure 2: A graphical illustration depicting the quantity of chosen articles categorized by year.

3.1 Arabic Dialect

Righi et al. [29], analyzed a political rumor case concerning the health of Algerian President Tebboune Abdelmadjid that took place between the end of 2020 and the beginning of 2021. They used the transfer learning approach and employed mBert, XLM-Roberta, and AraBERT models on a dataset of YouTube comments collected via the YouTube API v3. The data, which comprised 3,147 comments, was annotated into six categories: Support, Deny, Query, Comment, Positive Comment, and Negative Comment. The highest performance was achieved by the AraBERT model, with an F1-score of 53% compared to 45% and 38% for mBert and XLM-Roberta, respectively. The authors attributed the better performance of the AraBERT model to its training on NLP tasks such as sentiment analysis, and its training on a purely Arabic corpus with 24GB text and a vocabulary of 64K.

Bousri et al. [21], discussed the issue of infobesity, where people are overwhelmed with news and information via social media, and the resulting spread of rumors and misinformation. The authors propose a rumor detection approach for Algerian Arabizi that uses associations between rumors and reactions of social network users, as well as features representing the semantics of users' expression. They use an attention mechanism to compute the importance of these features, and test different classification models (LSTM, GRU, CNN) and textual representations (Word2vec, bag of n-grams, ELMo) to study

the associations between rumors and user reactions. They gatherd a data of 9.528 youtube comments regarding the death of the former Algerian president (abdelaziz boutaflika), and According to the results of the experiments, not all classifiers and representations benefited from the use of associations in improving the classification. However, the use of associations significantly enhanced the construction of the LTSM model for Word2vec and ngrams bag representations, showing an improvement of more than 10%. The effectiveness of associations in rumor detection is expected to become even more apparent with larger datasets.

Alorini and Rawat [11] explored the growing use of social media, particularly Twitter, in the Arab region for sharing news and spreading propaganda, including adult content and false political news. Despite the illegality of distributing adult materials within the Arab region, spammers continue to use these sites. they gathered 2000 tweets using twitter API, The authors of the paper examine user and content attributes to distinguish between legitimate and illegitimate users and use machine learning algorithms, including Naive Bayes and Support Vector Machine, to detect spam on Twitter. The results indicate that Naive Bayes is more accurate for detecting spam in Gulf Dialectical Arabic tweets achieving 92% F1-score while SVM got 83%.

3.2 Modern Standard Arabic (MSA)

Saadany et al. [51], analyzed the linguistic characteristics of Arabic fake news with satirical elements and developed machine learning models to identify it. The study used a dataset of 3185 articles obtained from two Arabic satirical news websites using a web scraping tool [2] and a dataset of 3710 articles from BBC and CNN news sources. The articles in both datasets dealt with political issues in the Middle East. The researchers achieved an accuracy of 98.6% using a CNN with pre-trained word embeddings. The study found that satirical fake news has distinct lexico-grammatical features that can be used to distinguish it from real news, even though it parodies real news.

Ali et al. [5], aimed to identify Arabic Twitter users who tend to spread fake news and limit the spread of misinformation. It created a dataset of 1,546 users, with 541 determined to be prone to spreading fake news based on their tweets and profile information. The study used linguistic, statistical, and profile features extracted from users' recent tweets to predict whether they were likely to spread fake news. Multiple learning models were applied and evaluated, with the logistic regression model achieving an F1 score of 73%. The approach was also tested on a benchmark English dataset and showed better results than the current state-of-the-art for this task.

Aljwari et al. [7], aimed to identify the factors contributing to the spread of Arabic fake news online through the use of machine learning models (Naive Bayes, Logistic Regression, and Random Forest) on publicly available news stories labeled as either FAKE or NOT-FAKE. A random partition was selected for testing and validation. Results showed the Random Forest Classifier had the highest accuracy 86% followed by Naive Bayes 84% and Logistic Regression 85%. The model found that the content's Term Frequency-Inverse Document Frequency (TF-IDF) features were crucial for Arabic fake news.

Thaher et al. [56], developed a smart fake news detection model for Arabic tweets on Twitter using NLP techniques, ML models, and Harris Hawks Optimizer for feature selection. The evaluation used 1862 [61] annotated tweets and tweepy[3] and employed the Bag of Words model with various term-weighting methods for feature extraction. 8 learning algorithms were tested with various combinations of user-profile, content-based, and word features. Logistic Regression with TF-IDF had the best performance with an 81.18% F1-score. The Harris Hawks Optimizer enhanced the learning model's performance. The proposed model was 5% more effective than previous methods on the same dataset.

Alotaibi and Alhammad [12] used a rule-based approach to classify Arabic fake news tweets related to Covid-19 into six categories: entertainment, health, politics, religious, social, and sports, and to analyze their spread. They used a text classification based on an Arabic dictionary on a dataset of 5 million tweets retrieved using hashtags. The model had an accuracy of 78.1%, with 70% precision and 98% recall, and detected over 26,006 fake news tweets. The study found that the number of fake news tweets decreased as Covid-19 awareness rose. The social category had the highest prevalence of fake news in most Arab

[2]Beautiful Soup, an open-source Python Library, was used in compiling the fake dataset.
[3]An easy-to-use Python library for accessing the Twitter API http://docs.tweepy.org/en/v3.5.0/index.html

Table 2: The summary of Related work (a).

Paper	Year	Language	Model	Dataset	Evaluation
Righi et al. [29]	2022	Algerian	mBert XLM-Roberta AraBERT	3,147 comments	F1 (mBert) = 45% F1 (XLM-Roberta) = 38% F1 (AraBERT)= 53%
Bousri et al. [21]	2023	Algerian	GRU LSTM CNN	9.528 comments	F1 (GRU) = 82% F1 (LSTM) = 82% F1 (CNN) = 67%
Alorini et al. [11]	2019	Gulf	NB SVM	2000 tweets	F1 (NB)= 92% F1 (SVM)= 83%
Saadany et al. [51]	2020	MSA	NB XGBoost CNN	3185 articles 3710 articles	F1 (NB) = 96.24% F1 (XGBoosT) = 96.81% F1 (CNN) = 98,49%
Ali et al. [5]	2022	MSA	RF XGBoost LR NN	1,949 true tweets 4493 false tweets	Best F1 (LR) = 73%
Aljwari et al. [7]	2022	MSA	RFC NB LR	The Ara News dataset	Accuracy (RF) = 86% Accuracy (NB) = 84% Accuracy (LR) =85%
Thaher et al. [56]	2021	MSA	LR with TF-IDF	1862tweets	F1 = 81.18%
Alotaibi et al. [12]	2022	MSA	Rule-based	5 milliontweets (related to Covid-19)	Accuracy = 78.1% Precision = 70% Recall = 98%
Shishah [54]	2022	MSA	JointBERT	Covid19Fakes Corpus (4954 Instances) Satirical Corpus (3242 Instances) AraNews Corpus (13398 Instances) ANS Corpus (4546 Instances)	F1 (Covid19Fakes) = 85% F1 (ANS) = 66% F1 (AraNews) = 80% F1 (Satirical) = 55%
Nagoudi et al. [46]	2020	MSA	mBERT, XLM-Rbase, XLM-Rlarge, AraBERT	AraNews + khouja dataset	Best F1 (XML-Rlarge) = 70.06%
Sawan et al. [52]	2021	MSA	LR		Accuracy = 82%
Alqurashi et al. [13]	2021	MSA	XGBoost	8786 Arabic COVIDtweets	AUC = 85.4%

countries except for Palestine, Qatar, Yemen, and Algeria, while fake news in the entertainment category had the least dissemination.

Shishah [54] proposed a BERT-based architecture for Arabic fake news detection using AI algorithms. The technique was evaluated on real Arabic fake news datasets and outperformed the current state-of-the-art model in two datasets (Covid19Fakes with 4954 instances and Satirical with 3242 instances). The proposed method had an average 10% improvement in F1 score across all datasets, with Covid19Fakes reaching 85%, ANS 66%, and Satirical 55% except for AraNews with 13398 instances which performed worse than AraBERT and QARiB. The results showed that the proposed method was effective and better than other baselines for detecting Arabic fake news.

Nagoudi et al. [46], proposed a method for generating Arabic manipulated (potentially fake) news stories using MADAMIRA [4] part-of-speech tagger and real news stories. The study also created the AraNews dataset, a large POS-tagged news corpus for future research, and conducted a human study to evaluate the impact of machine manipulation on text authenticity and human ability to detect manipulated Arabic text. The study introduced the first models for detecting manipulated Arabic news and achieved top results in Arabic fake news detection with a macro F1 score of 70.06%.

Sawan et al. [52], aimed to create a reliable method for identifying fake news in Arabic tweets from the dataset [61]. Due to limitations, the number of tweets was reduced to 1862, and the sentiments used were based on [32]. NLP techniques were used to structure the tweets, and recursive feature elimination was employed to eliminate less informative features. Advanced machine learning algorithms, including logistic regression, were used to build the prediction model. The results showed that logistic regression had the best performance and recursive feature elimination improved the overall accuracy, which was 82%.

Alqurashi et al. [13], address the problem of misinformation in Arabic Twitter content related to COVID-19. They collected a dataset of 8786 Arabic COVID tweets through TSAI[5] and tweepy and manually categorized them as containing misinformation or not. Traditional and deep machine learning models were applied to the data, utilizing features such as word embeddings and word frequency. The results showed that the Extreme Gradient Boosting (XGBoost) model with FASTTEXT word embeddings was the most accurate among the traditional classifiers, with an AUC of 85.4% in detecting COVID-19 misinformation on Twitter.

Bahurmuz et al.[17], propose and evaluate an Arabic rumor detection model using transformer-based deep learning architecture, specifically the AraBERT and*MARBERTv2* models, extensions of the BERT model. The model was tested on three Arabic datasets: COVID-19 misinformation [13], fake news detection [44], and Arabic rumor-non-rumor tweets [14] consisting of 36,308 tweets and achieved a maximum accuracy of 97%. The study also tackled the issue of imbalanced training datasets by using two sampling techniques. The proposed method was found to perform better than other existing Arabic rumor detection methods.

Nassif et al. [47], focused on detecting fake news in Arabic, an area that has received limited attention in previous research. They contribute a large and diverse Arabic fake news dataset and develop and evaluate transformer-based classifiers for identifying fake news using eight state-of-the-art Arabic contextualized embedding models, using two datasets: one translated from English and one gathered from Twitter. The models(Giga-Bert,Roberta-Base, AraBert, Arabic-BERT, ARBERT, MarBert, Araelectra and QaribBert)were found to be effective, where out of the 8 models we used, ARBERT and Arabic-bert performed the best with 98.8% and 98% respectivly, surpassing the translated data by 2 to 9 points. The study also provides a comprehensive analysis of these models and comparisons with other fake news detection systems.

Wotaifi and Dhannoon [60], aimed to build a better model for identifying fake news in Arabic by incorporating text, user features, and text features. The dataset used is based on the same dataset from [61], which includes 1862 tweets, but has been expanded using the same attributes. The tweet content is analyzed using the TF-IDF method, and a fuzzy model is used to identify relevant user features. The random forest algorithm is modified and improved, resulting in better performance compared to other machine learning methods such as Naive Bayesian and SVM. The accuracy of the Improved Random

[4]MADAMIRA was trained on the training sets of Penn Arabic Treebank corpus (parts 1, 2 and 3) [42] and the Egyptian Arabic Treebanks [43].

[5]Twitter streaming application interface

Forest method is 0.895, while the accuracy of the Naive Bayesian and SVM techniques are 80.9% and 84.8%, respectively.

Alazab et al. [2], presented a machine learning-based system for detecting fake news in Arabic. The system was trained on a dataset of 206,080 tweets from Twitter, which was collected using the API search. The algorithm employs TF-IDF to extract features from the dataset and ANOVA to select a subset of those features. Nine different machine learning classifiers were used to train the model, including Naive Bayes, KNN, SVM, Random Forest, J48, Logistic Regression, Random Committee, J-Rip, and Simple Logistics. The experiments showed that the Random Forest and Random Committee classifiers achieved the highest accuracy of 97.3%, with training times of 4403s and 525s, respectively.

Himdi et al. [30], focused on identifying fake news in the Arabic language using a machine learning model that classifies news sentences based on their credibility. The study introduced a first of its kind dataset of 700 real Arabic news sentences from 2013 to 2018 focusing on the topic of Hajj which had an inter-annotator agreement of 0.714 through Fleiss Kappa and created a fake news dataset through crowdsourcing. The authors use Arabic lexical wordlists and a NLP tool to extract textual features, POS tagging, syntactic tagging and emotions tagging as well from the articles and found that the proposed module outperforms human performance, achieving an accuracy of 78%.

Ahmed et al. [1], discussed the unprecedented amount of information being created and shared on social media platforms such as Facebook and Twitter, with some of it being misleading and irrelevant. Classifying text articles as misinformation or disinformation is a difficult task, even for domain experts who need to explore various aspects to determine their truthfulness. The authors propose an approach that uses an ensemble of machine learning algorithms DT, KNN, RF, LR, LSVM to classify news articles automatically. They explore different textual properties that can distinguish between fake and real content and train a combination of machine learning algorithms using various ensemble methods. The proposed approach is evaluated on four real-world datasets, the first one is 44,898 articles, the second one is 20,386 articles, the third one is 3,352 and the fourth one is a combined dataset of all three datasets totalling at 68636 articles, the experimental results show that the ensemble learner approach performs better than individual learners with up to 91% F1-score in the combined dataset.

Sorour and Abdelkader [55], presented a literature review on the detection of fake news (FN), focusing on Arabic FN (AFN) and the recent attention given to its detection. The authors propose an Arabic FN detection (AFND) system based on a hybrid deep learning (DL) model, specifically a combination of conventional neural network and long short-term memory (CNN-LSTM) modalities. The input dataset is preprocessed through discretization and normalization, and word vectors are included as Pre-trained vectors on Arabic news. The authors also introduce the JSO optimization algorithm to automatically determine the optimal structure for the CNN-LSTM. Experimental results show that the proposed CNN-LSTM outperforms other recent models with an accuracy of 81.6%. Overall, the authors demonstrate the potential of their proposed methodology in improving the detection of AFN.

Alzanin et al. [15], discussed the importance of automated classification of tweets due to the rapid growth in the number of tweets being published daily on Twitter. The paper proposes a scheme to classify Arabic tweets based on their linguistic characteristics and content into five different categories using word embedding using Word2vec and stemmed text with term frequency-inverse document frequency (tf-idf) representations. The authors collected and manually annotated a dataset of approximately 35,600 Arabic tweets and tested three different classifiers: Support Vector Machine (SVM), Gaussian Naive Bayes (GNB), and Random Forest (RF). The results showed that the RF and SVM with radial basis function (RBF) kernel performed equally well when used with stemming and tf-idf, achieving macro-F1 scores ranging between 98.09% and 98.14%. However, the GNB with word embedding performed poorly. The proposed approach outperformed the current state-of-the-art score of 92.95% using a deep learning approach, RNN-GRU (recurrent neural network-gated recurrent unit).

Albalawi et al. [3], discussed the proliferation of rumors on social media platforms and the need for artificial intelligence techniques to detect them. Most existing works in Arabic language focus on textual features of tweets, but tweets contain different types of content, including visual features that play an essential role in rumor diffusion, they used arafacts which has 6,222 claims which 4141 of them are videos or images and from that they extracted 1726 tweets. The study proposes an Arabic rumor detection model that combines textual and visual image features through early and late fusion. The authors conducted experiments to select the best feature extractors and found that the effectiveness of

textual features is higher than that of multimodal models.*MARBERTv2* achieved the highest result with an F1 score of 90%. *MARBERTv2* and an ensemble of VGG-19 and ResNet50 were used as feature extractors for building a multimodal model, and the results were compared with those of single models. The paper concludes that the proposed model is effective in detecting rumors on Twitter, and the textual features are more effective in this task than multimodal models.

3.3 Datasets

Bsoul et al. [23], presented a new Arabic clickbait news dataset, aimed at classifying news headlines as Clickbait or Not Clickbait with machine learning models. The dataset was created by sampling over 3000 news records from tweets of 24 Jordanian news publishers over five months, labeled by three annotators with 18% classified as click bait and 81% receiving unanimous agreement. The dataset was evaluated with machine learning models such as Logistic Regression, SVM, Random Forest, Naive Bayes, SGD, Nearest Neighbor, and Decision Tree, achieving Macro F1-Score values up to 81%, indicating the possibility of detecting clickbait news headlines automatically with machine learning.

Khalil et al. [35], created the first extensive Arabic fake news corpus (AFDN), with 606912 articles gathered from 134 online news sources and fact-checked by an Arabic platform. The study evaluated the effectiveness of various machine learning algorithms in detecting fake news in the corpus. Deep learning models were found to perform better than traditional machine learning models, but the results also showed that the corpus was challenging due to noise, leading to underfitting and overfitting problems for models trained on it, never the less capsule network model was the highest performing of the banch reahcing 71%.

Mohdeb et al. [45], explored the issue of fake news concerning Covid-19 in Arabic content. The authors have created a unique dataset of Arabic fake and true news from trustworthy sources, specifically focusing on Covid-19 misinformation in Arabic. The dataset consists of 1280 articles labeled into 12 categories. It was used to evaluate the performance of baseline classification models, with results showing high performance and the SVM classifier demonstrating exceptional accuracy at 94%.

Alkhair et al. [8], examines the issue of fake news in the Arabic world by analyzing content on YouTube. The authors present a new corpus for fake news analysis in Arabic, consisting of 3435 comments after cleaning and focusing on commonly rumored topics. The authors provide details about the corpus and data collection process, conduct an exploratory analysis of the data to understand the spread of rumors, and evaluate the performance of three machine learning classifiers (SVM, DT, and MNB) in identifying comments containing rumors. The SVM classifier showed the highest accuracy, reaching 95.35% when categories were combined.

Chouigui et al. [27], introduced a new online Arabic corpus of news sentences called ANT Corpus, which was gathered from RSS Feeds. The authors used this corpus for Text Classification (TC) by applying the SVM and Naive Bayes classifiers to classify articles into their respective predefined categories. The study also examined the impact of terms weighting, stop-words removal, and light stemming on Arabic TC, SVM got 82.77% when using a smal subset from their data which was balanced and containing the first 172 article from every category. Results showed that text length significantly impacts TC accuracy and that titles are not adequate for good classification rates. The SVM method was found to provide the best results for classifying both the titles and texts

Alderywsh et al. [4], aimed to tackle the issue of fake news spreading in Arab societies through advanced technologies, such as machine-generated text. The study employed an iterative method to develop a fake news detection system that uses machine learning algorithms in Python to distinguish between genuine and fake news in the Arab context. The system, called Tebyan, allows users to assess the credibility of news in Arabic newspapers. Results showed that the Support Vector Machine classifier and the linear Support Vector Classification algorithm exhibited high performance, with an average of 90% accuracy in detecting fake news. The study concludes that using machine learning algorithms with Python programming language is a fast and effective approach to evaluating the credibility of news in Arab societies.

Ameur et al. [16], discusses the emergence of false and misleading information during the COVID-19 pandemic, which has complicated response efforts. Social media platforms have contributed to the spread of rumors, conspiracy theories, and prejudice. To combat the spread of fake news, researchers have

Table 3: The summary of Related work (b).

Paper	Year	Language	Model	Dataset	Evaluation
Bahurmuz et al. [17]	2022	MSA	AraBERT MARBERTv2	COVID-19 misin-formation Fake NewsDetection Arabic rumor-non-rumor tweets	Best F1AraBERT = 93% (Arabic rumor-non-rumor tweets) Best F1 MARBERTV2 = 96% (Arabic rumor-non-rumor tweets)
Nassif et al. [47]	2022	MSA	Giga-Bert, Roberta-Base AraBert, Arabic-BERT ARBERT, MarBert Araelectra and QaribBert	16K instances (Translated from English) and 10K instances (collected)	Best F1 (ARBERT) = 98.9% Best F1 (GigaBert-base) = 91.8%
Wotaifi et al. [60]	2022	MSA	RF NB SVM	1862 tweets	Accuracy RF = 89.5% Accuracy SVM = 84.5% Accuracy NB = 80.5%
Alazab et al. [2]	2022	MSA	NB, KNN, SVM, RF, J48 LR RC, J-Rip SL	206,080 tweets	Best Accuracy (RF and RC) = 97.3 %
Himdi et al. [30]	2022	MSA	NB RF SVM	Parallel Corpus (700 real / fake sentences)	Best Accuracy (RF) = 79%
Ahmed et al. [1]	2020	MSA	LR LSVM MLP KNN	DS1 = 44,898 articles DS2 = 20,386 article DS3 = 68636 articles	F1 (LR) = 91%
Sorour et al. [55]	2022	MSA	CNN-LSTM	ANS Corpus	Accuracy = 81.6%
Alzanin et al. [15]	2022	MSA	SVM GNB RF	5,600 tweets	F1 (SVM) = 98.14% F1 (GNB) = 92.95% F1 (RF) = 98.09%
Albalawi et al. [3]	2023	MSA	MARBERTv2	1726 tweets	F1 = 90%

made significant efforts to build and share COVID-19 related datasets. The paper presents a manually annotated Arabic COVID-19 fake news and hate speech detection dataset called AraCOVID19-MFH, which contains 10,828 tweets with 10 different labels, he found that using finetuning with arabert, the data got 94.17% while got 95.78% while using arabert cov19 pretrained model.

Al Zaatari et al [61] detail the process of developing a corpus for credibility analysis and demonstrate their utility. The corpus comprises 2,708 Arabic tweets, each manually labeled as either credible or non-credible, the corpus was tested with CAT classifier in the combined task and it got F1-score of 70.1%.

Table 4: The summary of Related work (Dataset).

Ppaer	**Year**	**Dataset**	**Best Performing paper**
[23]	2022	3000 news records	[23]
[35]	2022	AFDN	[35]
[45]	2021	1280 articles	[45]
[8]	2019	3435 YouTube comments	[8]
[27]	2017	ANT Corpus	[27]
[4]	2021	Tebyan	[4]
[16]	2021	AraCOVID19-MFH	[16]
[61]	2016	2,708 Arabic tweets	[60]

4 Model architecture

Few-Shot learning techniques are popular solutions for situations with limited labeled data. They involve adapting pre-trained language models for specific tasks with minimal training examples. Approaches like in-context learning (ICL) as used in GPT-3 [22], task specific prompt learning [53] and adaptation approach which performs a localized updates that concentrate changes in a small set of model parameters like using adapters [31]; [48];[58], demonstrate good efficiency but they may not always be practical for all researchers and industries due to: (i) requirement of billion parameter language models; (ii) subject to high variability due to their reliance on manually crafted prompts; (iii) the deployment of those models is critical in terms of infrastructure and cost paying. To overcome these limitations, a new framework called *SetFIT* (Sentence Transformer Fine Tuning) [57] has been proposed.
SetFIT is an efficient and prompt-free framework for Few-Shot fine-tuning of Sentence Transformers (ST) [49]. It works by first fine-tuning a Pre-trained ST on a small set of text pairs in a contrastive Siamese manner. The model then generates rich text embeddings, which are used to train a classification head as shown in figure 3. This simple framework does not require prompts or verbalizers, and achieves high accuracy with orders of magnitude fewer parameters than existing techniques.

SetFIT employs ST, a distinct type of transformer model, which requires an understanding of its training and design. Using Siamese networks and triplet network structures, Sentence Transformers adapt pre-trained transformer models to produce semantically meaningful sentence embeddings. Unlike standard transformer models that generate token embeddings, Sentence Transformers create embeddings for entire sentences. These embeddings are semantically informed, meaning that semantically similar sentences have minimized distances, while semantically distant sentences have maximized distances. As a result, Sentence Transformers provide dense vector representations for textual data. Notably, the architecture of standard transformer-based models, such as BERT [28] and RoBERTa [41], presents challenges in utilizing them for semantic similarity search and unsupervised tasks like clustering. Sentence Transformers address this issue by leveraging the semantic properties of sentences more effectively. As demonstrated in the figure 1, the training of *SetFIT* exhibits a simple two-step: sentence transformer fine-tuning and classification head training.

ST fine-tuning Due to the data limitation, the first step of the Fine-tune ST is based on a contrastive training approach. This approach, commonly used for image similarity [38], involves generating positive and negative triplets based on a small set of labeled examples D = (xi, yi), where xi and yi represent

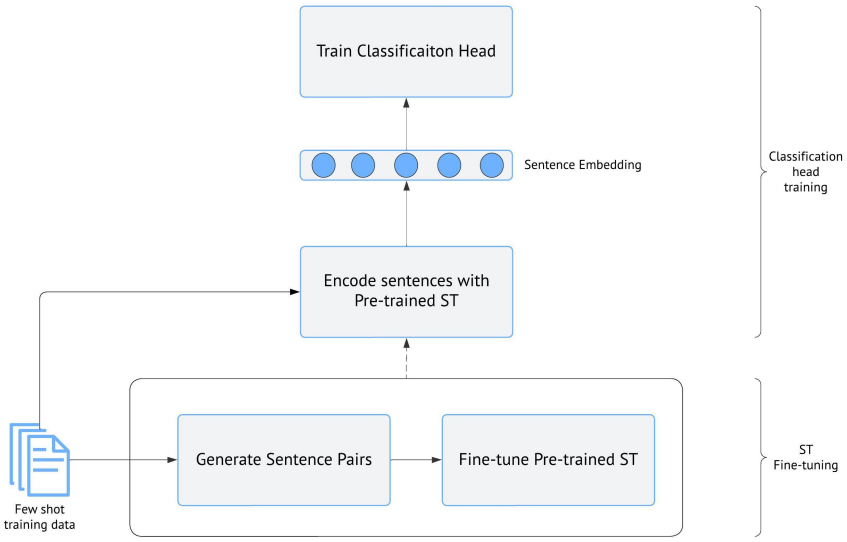


Figure 3: A visual representation of *SetFIT*'s fine-tuning and training process.

sentences and their respective class labels. For each class label c in C, a set of R positive triplets Tc_p and a set of R negative triplets Tc_n is created. Positive triplets consist of pairs from the same class, while negative triplets contain a sentence from one class and another from a different class. In each sentence-pairs iteration, we generate 2xK training pairs, where K is the total number of training samples in the task. The contrastive fine-tuning dataset generated T is formed by concatenating the positive and negative triplets across all class labels. After this, a ST model is Fine-tuned on the T dataset (or triplets) to generate an adapted ST.

Classification head training The second step consists of using the original (limited) training data and generating sentence embeddings from the adapted ST. These embeddings (as inputs) and the original class labels are now used as the dataset to train the classification head. In the original research paper [49], a logistic regression (LR) model is used as the text classification head. For the inference phase, each test sentence is encoded with the adapted ST, and then its category is predicted by the LR model.

5 Experimental Setup

5.1 Dataset

In this paper, we utilised three different sources for our dialectal algerian dataset, which include YouTube comments, handmade sentences, and manually translated sentences from the Khouja dataset[36] by native Algerian speakers.

Because the Algerian dialect is not merely a derivative of MSA, it has evolved into a distinct language through a combination of Arabic, Turkish, French, English, Spanish, and MSA influences. This unique linguistic blend makes it challenging for even MSA speakers to comprehend and virtually impossible for others. Consequently, we specifically opted to focus on YouTube comments related to Algeria, allowing locals to employ their Algerian dialect. We then selectively chose sentences or comments containing the Algerian dialect, excluding those in MSA or any other language. Our criteria for selection were centered on sentences that conveyed specific meaning and had a newsworthy aspect, rather than including just any comment.

a user is deemed trustworthy based on the current news and trusted sources on that period of time, as for the translated part it was already fact checked by the author of the paper since we were just changing the language of the news from MSA to AD. We used some credibility criterias like selecting specific subjects in which we know the true from the fake and then we gather the users opinion in form if short news

sentences, as well as relying on the timeline of the event since whats true now is not always true even tho there are some events that would never change like a football team wining the worldcup in a specific year.
In this study, we constructed a dataset for the Algerian dialect from three distinct sources: YouTube comments, original sentences, and sentences manually translated by native Algerian speakers from the Khouja dataset [36]. We specifically focused on YouTube comments related to Algeria, where locals naturally employ the Algerian dialect. We selected sentences and comments exclusively in the Algerian dialect, omitting those in MSA or any other language. Our selection criteria prioritized sentences conveying specific meaning and having newsworthy relevance rather than including arbitrary comments.
In ensuring the trustworthiness of user-generated content, we evaluated the credibility of users based on current news and trusted sources at the time. For translated content, it underwent thorough fact-checking by the author to ensure accurate language conversion from MSA to the Algerian dialect. Additionally, we considered the timeline of events, recognizing that the truth can evolve over time, although some events, like a football team winning the World Cup in a specific year, remain constant.
The dataset consists of 5905 sentences, which were annotated based on their credibility and realness. There are seven categories of the dataset as shown in Table 5, which includes accidents in Algeria (a big portion was gathered from annually accidents report published by the minister of internal), car prices in Algeria, COVID-19, tourism in Algeria, , eCommerce in Algeria, politics, and the qualification of Algeria to the 2022 World Cup held. The dataset is almost balanced, consisting of 3013 real news and 2892 fake news sentences . The labels were applied based on a thorough analysis of the credibility and realness of the news.

Table 5: Examples of categories within the dataset.

Category	**Algerian text**	**English translation**
Car accident	٧ مجروحين بسباب اكسيدو في لغواط	7 people injured due to a car accident in laghouat
Car price	دخلو طونوبيلات جدد و مطاحش سوق	New cars arrived and the price still the same
COVID-19	جابو دوا تاع كورنا وبدات تتروح	they brought the corona' vaccine and the virus start to disappear
Tourism	السياحة كاينا فدزاير	Tourism exists in Algeria
eCommerce	خطأ نعرف ناس بزاف بداو ايكومرس بزيرو دراهم	Wrong, I know lot of people started ecommerce with zero budget
Politics	سومة البترول طلعت قبل ما يضربو سوريا	The price of oil increased before the attack on Syria
World Cup	فشلت الجزائر في التأهل إلى كأس العالم	Algeria failed to reach the World Cup

Overall, this dataset provides a comprehensive and diverse set of news categories and sources, allowing for effective training and testing of machine learning models for the detection of fake news. It is also important to note that ethical considerations were taken into account during the data collection process, with all sensitive information anonymized and user consent obtained where necessary. note that algerian dialect is a very low resource dialect to gather data for, algerians are not that open to twitter and other big social media apart from facebook thus the news and publications are all made in MSA which made it extremely hard to gather data on the algerian dialect.

5.2 Models Description

The performance of SetFit is influenced by various factors, including the type of the ST, input data selection based on features, and the selection of hyperparameters. The Hugging Face Transformers library provides access to the model hub, which enables the use of any ST. In our study, no STs were

available for MSA or Dialectical Arabic, and therefore, we used regular models that were trained using Masked Language and Next Sentence Prediction objectives. Despite the lack of explicit guidelines for choosing the best model for a downstream NLP task, some general rules of thumb exist, particularly for similarity tasks. As the Fake News task involves text classification, we hypothesized that an embedding model that has achieved high performance in different classification tasks would be advantageous.
For that, we compared the performance of multiple Arabic pre-trained models trained on Arabic dialects, namely *MARBERTv2*, *DziriBERT* and *DarijaBERT* and models that were mainly trained on MSA which are *ARBERT* and *CAMeLBERT* and models that are for MSA but contain dialects in them such as *AraBART* and *AraBERTv2* in order to figure the most suitable one for the Algerian fake news task and lead to achieve good results. This section provides an overview of the pre-trained models utilized, taking into account different factors such as vocabulary size, training data source, parameter count, and the languages incorporated during the training phase. Table 6 presents a comparison of the pre-trained models employed in this study.

Table 6: Summary of the used Arabic pre-trained models in terms of dataset, size, vocabulary size, and training language.

Model	**Size(Params)**	**DataSet**	**Vocab size**	**Language**
MARBERTv2	163M	Arabic Twitter	100K	Dialect
ARBERT	163M	Several (6 sources)	100K	MSA
DziriBERT	124M	Twitter	50K	Dialect
AraBERTv2-base	136M	OSCAR, OCIAN, alssafir corpus	64K	Dialect/MSA
DarijaBERT	100M	40 different Morrocan channels	80K	Dialect
AraBART	139M	20GB corpus	50K	Dialect/MSA
CAMeLBERT_MSA	110M	Arabic Gigaword, El-Khair, OSIAN, Wikipedia Corpus	30K	MSA

5.3 Hyper-Parameter Setting

For our experiments, we utilized version 4.9.2 of the Hugging Face Transformers library, which offers a flexible and user-friendly platform for training and deploying cutting-edge NLP models. To compute our experiments, we employed Google Colab Pro, which provides access to a high-performance GPU V100 and the convenience of running Python code in a Jupyter notebook, as well as the automatic storage of data and scripts to Google Drive. Additionally, we utilized Kaggle, which offers a variety of resources, including GPU and TPU support for training machine learning models, and in our case we utilized the P100 GPU. The experiments had a total run time of up to 9 hours, divided into 5 epochs of 1.8 hours each, with Kaggle being the primary computing resource. Although Colab Pro has a larger memory pool, it only provides 100 compute units per 10 dollars. Furthermore, all hyper-parameters for the models used in our experiments are presented in the GitHub repository[6] and Table 7.

[6]https://github.com/amincoding/Algerian_fake_news

Table 7: Hyper-parametres summary.

number of samples	number of iterations	batch size	epochs	learning rate
400	80	16	5	1 exp −3
800	80	32	5	1 exp −3
1500	80	32	5	1 exp −3

The choice of the training sample sizes was determined following a rigorous testing process, in accordance with the methodology outlined in the SetFit research paper. These sample sizes were identified as optimal, taking into account the limitations of our initial dataset. Efforts to incorporate a larger number of samples were constrained by resource limitations and were therefore not pursued further.

6 Results and Discussion

This study encompasses four discrete experiments, as introduced in the 1 section. All outcomes were quantified using key metrics encompassing Accuracy, Precision, Recall, and F1-Score. Each subsequent sub-section provides a comprehensive summarization of the research questions and their results through the presentation of tables and figures, which are subsequently followed by an in-depth discussion segment.

6.1 Impact of training sample variation on SetFit performance

The primary objective of this experimental investigation is to assess how the quantity of training samples per class influences the performance of various pre-trained Dialectical Arabic models. In this study, we will analyze the effectiveness of five key models: *MARBERTv2*, *DziriBERT*, *AraBART*, *AraBERTv2*, and *DarijaBERT*, while varying the number of training samples per class, as outlined in Table 7.
Additionally, we will compare the BERT and BART architectures for sentence classification tasks. These experiments aim to identify the optimal number of training samples per class for each model and to evaluate their performance based on metrics such as accuracy, precision, recall, and F1-score. In essence, this study aims to provide valuable insights into the impact of training sample size on Few-Shot learning performance, leveraging the *SetFIT* framework.

The experimental outcomes are summarized in Table 8. Notably, *DziriBERT* outperformed other pre-trained models with 400 training samples, achieving an F1-score of 0.6611, which surpassed other models by a 3% margin. This can be attributed to its monolingual training data tailored specifically to the Algerian dialect. These results suggest that a smaller dataset and a mono-dialect model, fine-tuned with contrastive learning, outperform larger multi-dialect models. With 800 training samples, *AraBERTv2* achieved good results, with *MARBERTv2* following closely with F1-scores of 0.6913 and 0.6707, respectively. Increasing the dataset size during training, benefited models with larger vocabularies and more pre-training data exposure. However, smaller models like *DziriBERT* experienced a decline in performance due to out-of-vocabulary words stemming from context expansion. In the case of 1500 training samples, both *MARBERTv2* and *DziriBERT* achieved the highest F1-scores of 0.7105 and 0.6897, respectively. This can be attributed to the substantial amount of Algerian dialect data saw during pre-training and fine-tuning. Notably, *AraBERTv2* achieved an exceptional recall rate, nearly 100%, indicating its ability to predict almost 99% of real cases, with a precision score of 0.5146, indicating some false positives. It's worth noting that varying sample sizes led to fluctuations in the results, as depicted in Figure 4. These variations underscore the sensitivity of model performance to the size of the training dataset.

Table 8: SetFIT framework's performance score for five pre-trained Arabic models for three different training set sizes, where 'N' represents the number of training examples per class. The highest score obtained in the training set evaluation is indicated in bold.

Model	Accuracy	Precision	Recall	F1 score
		N = 400		
MARBERTv2	0.6282	0.6366	0.6334	0.6350
AraBART	0.6113	0.6074	0.6749	0.6394
DarijaBERT	0.6011	0.5940	0.6915	0.6390
DziriBERT	0.6164	0.6021	**0.7330**	**0.6611**
AraBERTv2	0.6291	0.6486	0.5970	0.6217
		N = 800		
MARBERTv2	0.6401	0.6293	0.7180	0.6707
AraBART	0.6418	0.6730	0.5804	0.6233
DarijaBERT	0.6452	0.6389	0.7014	0.6687
DziriBERT	0.6206	0.6465	0.5671	0.6042
AraBERTv2	0.6604	0.6451	0.7446	**0.6913**
		N = 1500		
MARBERTv2	0.6909	0.6808	0.7429	**0.7105**
AraBART	0.6748	0.6809	0.6832	0.6821
DarijaBERT	0.6638	0.6886	0.6235	0.6544
DziriBERT	0.6663	0.6566	0.7263	0.6897
AraBERTv2	0.5182	0.5146	**0.9917**	0.6776

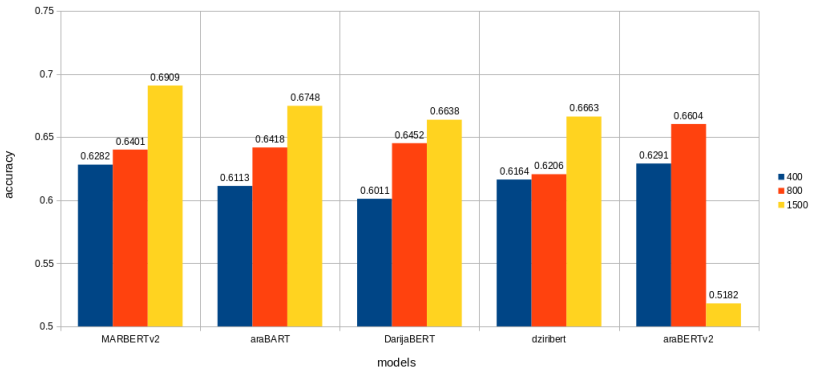


Figure 4: Visualization of the models' accuracy which was assessed across three different sample sizes per class.

6.2 Impact of normalization from Algerian Dialect to MSA

In this subsequent experiment, we sought to evaluate the effectiveness of the same Arabic pre-trained models as those utilized in the previous experiment, with the added step of normalizing the dataset using a pre-trained translator model designed specifically for translating Algerian dialect text to MSA version. This normalization was introduced to ensure fairness in performance evaluation.

For this experiment, we maintained a consistent training set size of 1500 samples per class, a decision based on the insights gained from the initial experiment. The pre-trained translator model employed here is based on the BART architecture, leveraging the *AraBART* model as a starting point. It was further fine-tuned using 13k parallel sentences, resulting in a respectable BLUE score of 29% during evaluation. Before incorporating this translator model into our experiment, we subjected it to a battery of tests, which verified its proficiency in maintaining both syntactical and contextual accuracy.

Table 9: The performance score of *SetFIT* framework on a normalized training set of 1500 samples per class.

Model	Accuracy	Precision	Recall	F1 score
		N = 1500		
MARBERTv2	0.6594	0.6632	0.6761	0.6693
AraBART	0.6816	0.6839	0.6998	0.6918
DarijaBERT	0.6676	0.6642	0.7035	0.6830
DziriBERT	0.6724	0.6780	0.6791	0.6798
AraBERTv2	0.6850	0.6820	0.7190	**0.7004**

Table 9 and Figure 5 summarize the results obtained from the experiment conducted on a normalized dataset. The results indicate that *AraBERTv2* outperformed the other models, achieving an F1-score of 0.7004, closely followed by *AraBART* with a score of 0.6918. Importantly, both of these models were pre-trained on the same dataset, highlighting that this dataset contains a notably higher proportion of MSA text compared to dialectal text.

In contrast, *MARBERTv2* achieved lower scores than the other models, primarily due to its pre-training dataset being exclusively focused on Maghreb's dialects. In comparison to the results from the previous experiment, these findings suggest that the normalization pre-processing step has a positive impact on models that saw MSA text during pre-training phase. However, it has a negative impact on models that were primarily trained on Arabic dialects, leading to an increased occurrence of out-of-vocabulary words.

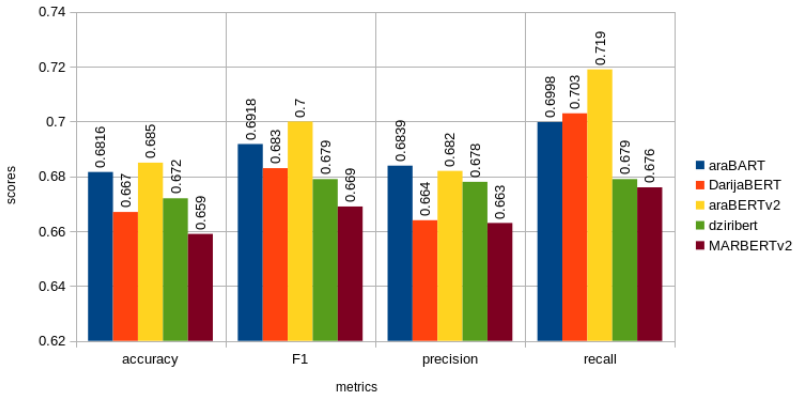


Figure 5: Visualizing the performance of the SetFIT framework across various pre-trained models on a normalized dataset with 1500 samples per class.

6.3 Cross-examining MSA-based models on the Algerian Dialect

This experiment aims to test the hypothesis that fully MSA-trained models, specifically *ARBERT* and *CAMeLBERT MSA*, can perform as well as or better than dialectal models. Table 11 provides an overview of the results from this experiment, outlining the performance of the models under both normalized and non-normalized conditions. The data illustrates that *CAMeLBERT MSA* surpassed *ARBERT* in both scenarios, achieving F1-scores of 0.6995 and 0.6812 without and with normalization, respectively. These results confirm the beneficial effect of normalization on models trained on MSA text, as observed in the previous experiment.

Table 10: The performance score of *SetFIT* framework for two Arabic pre-trained models on a normalized and non normalized training set of 1500 samples per class.

Model	Accuracy	Precision	Recall	F1 score
N = 1500, Normalized = NON				
ARBERT	0.6579	0.7068	0.5638	0.6273
CAMeLBERT_MSA	0.6960	0.7060	0.6932	**0.6995**
N = 1500, Normalized = YES				
ARBERT	0.6686	0.6742	0.6748	0.6744
CAMeLBERT_MSA	0.6672	0.6666	0.6965	**0.6812**

However, it is important to note that *CAMeLBERT MSA* experienced a decrease in performance when utilizing the normalized dataset. This decline can be attributed to the emergence of out-of-vocabulary words that were not encountered during the training phase, indicating that domain-specific terminology was absent from the training set. In contrast, *ARBERT* benefited from training on six distinct sources of text, enabling it to effectively handle a broader range of words and thereby improving its performance across diverse text domains. Figure 6 provides a visual comparison of both models' performance on an unnormalized dataset.

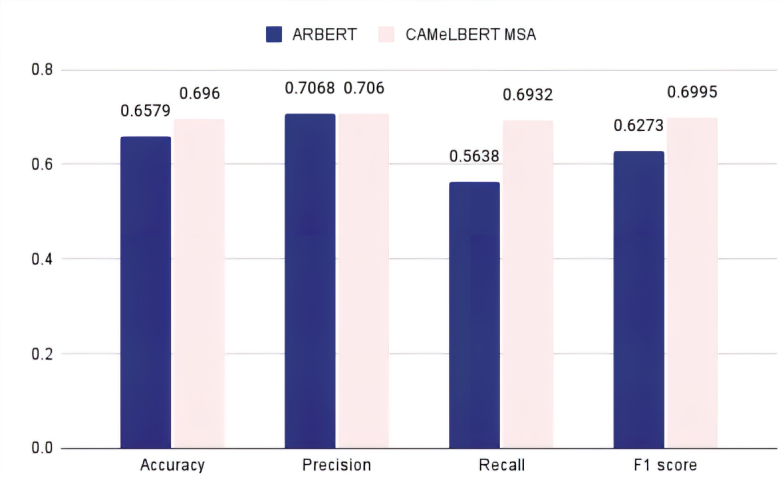


Figure 6: Visualizing the performance of models trained solely on MSA data in the context of dialectal testing data.

6.4 Comparing standard fine-tuning and Few-Shot learning approaches

In the last experiment, we carried out a comparative assessment of the performance of a transformer-based model fine-tuned with 1500 samples per class using the standard approach and the Few-Shot learning framework (SetFIT) trained on the same amount of data. This analysis aimed to identify the more efficient approach for dealing with low-resource data. The outcomes of this experiment are succinctly summarized in Table 11.

Table 11: Comparaison between the performance score of *SetFIT* framework and standard fine-tuning using the same Arabic pre-trained model and the training set of 1500 samples per class.

Model	Accuracy	Precision	Recall	F1 score
MARBERTv2 (SetFIT)	0.6909	0.6808	0.7429	**0.7105**
MARBERTv2 (Fine-tuned)	0.64	0.64	0.65	0.65

The results unequivocally indicate that *SetFIT* surpassed the performance of the fine-tuned model across all metrics, underscoring Few-Shot learning as the preferred approach for low-resource languages.

These findings affirm the efficacy of pre-trained language models in fake news detection, with *MARBERTv2* emerging as the most promising candidate, thanks to its extensive Arabic dialect dataset, primarily centered on Algerian dialects, and its substantial quantity.

Additionally, the evaluation of these models using various classifiers provides valuable insights into their adaptability to different scenarios, enhancing their practical applicability. For a thorough visual representation, Figure 7 illustrates a comprehensive comparison of both approaches across all metrics.

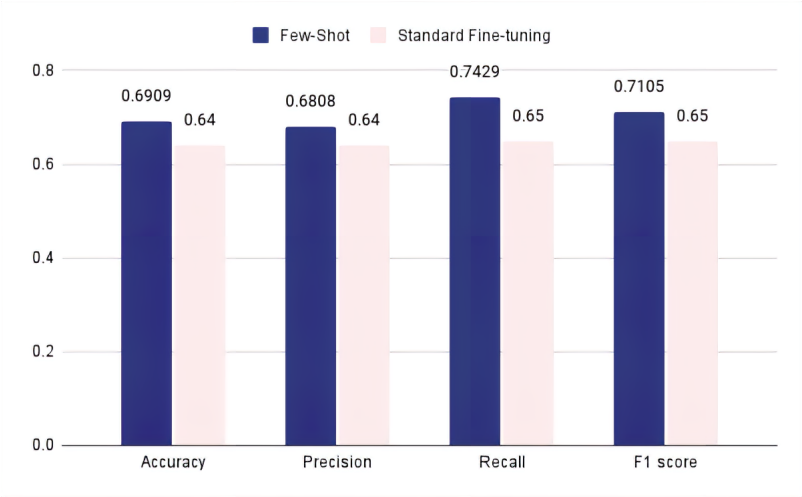


Figure 7: Visual comparison between the Few-Shot learning approach and the standard fine-tuning method using *MARBERTv2* on a dataset of 1500 samples per class.

7 Conclusion

In conclusion, this study has introduced a Few-Shot-based model for the detection of fake news in low-resource languages, with a specific focus on the Algerian dialect. We compared multiple Arabic pre-trained models based on various factors such as vocabulary size, training data source, parameter count, and the inclusion of languages during the training phase. Our performance evaluation criteria included accuracy, precision, recall, and F1-score.

The results have highlighted that the *MARBERTv2* model stands out as the top-performing model for detecting fake news in the Algerian dialect, achieving an impressive F1-Score of **0.7105**. Notably, *AraBERTv2* achieved the highest Recall of **0.9917**, demonstrating its ability to effectively identify true cases as true. All of this was accomplished using a dataset comprising 1500 samples from our corpus, and the Few-Shot-based model (SetFIT) consistently outperformed fine-tuning when utilizing the same dataset, reaffirming the potency of SetFIT in low-resource data scenarios.

Furthermore, with regard to the normalization pre-processing and the utilization of fully Modern Standard Arabic (MSA)-trained models such as *ARBERT* and *CAMeLBERT MSA*, these models did not perform as effectively as their dialectal counterparts. This study makes a valuable contribution to the advancement of Natural Language Processing (NLP) models for low-resource languages and underscores the significance of employing dialectal models for fake news detection in dialectal languages.

Based on the results obtained in this study, several potential avenues for future research can be explored:

- One potential area of exploration could be to expand the dataset with more news and other Algerian sub-dialects.

- Another possible direction could be to explore the impact of using other pre-processing techniques such as segmentation in order to reduce OOV.

- Additionally, investigating the use of multi-task models could potentially further improve the performance by adding more information such as sentiment analysis.
- Finally, exploring the effectiveness of using transfer learning techniques to adapt pre-trained models to specific low-resource languages could be another fruitful area for future research.

References

[1] Iftikhar Ahmad, Muhammad Yousaf, Suhail Yousaf, and Muhammad Ovais Ahmad. Fake news detection using machine learning ensemble methods. *Complexity*, 2020.

[2] Moutaz Alazab, Albara Awajan, Ammar Alazab, Abeer Alhyari, and Reem Saadeh. Fake-news detection system using machine-learning algorithms for arabic-language content. *Journal of Theoretical and Applied Information Technology*, 100(16):5056–5069, 2022.

[3] Rasha M. Albalawi, A. Jamal, A. Khadidos, and Areej M. Alhothali. Multimodal arabic rumors detection. *IEEE Access*, 2023.

[4] Lamya Alderywsh, Aseel Aldawood, Ashwag Alasmari, Ashwag Alasmari, Farah Aldeijy, Ghadah Alqubisy, and Sarah Alawwad. Tebyan: Fake news detection system (preprint). *null*, 2021.

[5] Zien Sheikh Ali, Abdulaziz AlAli, and Tamer Elsayed. Detecting users prone to spread fake news on arabic twitter. *OSACT*, 2022.

[6] A. Aliwy, Hawraa A. Taher, and Zena AboAltaheen. Arabic dialects identification for all arabic countries. *Workshop on Arabic Natural Language Processing*, 2020.

[7] Fatima Aljwari, Wahaj Alkaberi, Areej Alshutayri, Eman Aldhahri, Nahla Aljojo, and Omar Abouola. Multi-scale machine learning prediction of the spread of arabic online fake news. *Postmodern Openings*, 2022.

[8] Maysoon Alkhair, Karima Meftouh, Kamel Sma-li, Kamel Sma-li, and Nouha Othman. An arabic corpus of fake news: Collection, analysis and classification. *International Colloquium on Automata, Languages and Programming*, 2019.

[9] Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2):211–36, May 2017.

[10] Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2):211–236, 2017.

[11] Dema Alorini and Danda B. Rawat. Automatic spam detection on gulf dialectical arabic tweets. *International Conference on Computing, Networking and Communications*, 2019.

[12] Fatimah L. Alotaibi and Muna M. Alhammad. Using a rule-based model to detect arabic fake news propagation during covid-19. *International Journal of Advanced Computer Science and Applications*, 13(1), 2022.

[13] Sarah Alqurashi, Batool Hamawi, Abdulaziz Alashaikh, Ahmad Alhindi, and Eisa Alanazi. Eating garlic prevents covid-19 infection: Detecting misinformation on the arabic content of twitter. 01 2021.

[14] Samah Alzanin and Aqil Azmi. Rumor detection in arabic tweets using semi-supervised and unsupervised expectationmaximization. *Knowledge-Based Systems*, 185:104945, 08 2019.

[15] Samah M. Alzanin, Aqil M. Azmi, and Hatim A. Aboalsamh. Short text classification for arabic social media tweets. *Journal of King Saud University - Computer and Information Sciences*, 2022.

[16] Mohamed Seghir Hadj Ameur, Hassina Aliane, and Hassina Aliane. Aracovid19-mfh: Arabic covid-19 multi-label fake news & hate speech detection dataset. *Procedia Computer Science*, 2021.

[17] Naelah O. Bahurmuz, Ghada A. Amoudi, Fatmah A. Baothman, Amani T. Jamal, Hanan S. Alghamdi, and Areej M. Alhothali. Arabic rumor detection using contextual deep bidirectional language modeling. *IEEE Access*, 10:114907–114918, 2022.

[18] A. Barhoumi. Une approche neuronale pour l'analyse d'opinions en arabe. (neural approach for arabic sentiment analysis). *null*, 2020.

[19] Dan Berkowitz and David Asa Schwartz. Miley, cnn and the onion. *Journalism Practice*, 10(1):1–17, 2016.

[20] Siham Boulaknadel. Traitement automatique des langues et recherche dinformation en langue arabe dans un domaine de spcialit : Apport des connaissances morphologiques et syntaxiques pour lindexation. *null*, 2008.

[21] Mohamed Charafeddine Bousri, Riad Bensalem, Samah Bessa, Zineb Lamri, Chahnez Zakaria, and Nabila Bousbia. Rumor detection inalgerian arabizi based ondeep learning andassociations. In Salim Chikhi, Gregorio Diaz-Descalzo, Abdelmalek Amine, Allaoua Chaoui, Djamel Eddine Saidouni, and Mohamed Khireddine Kholladi, editors, *Modelling and Implementation of Complex Systems*, pages 165–176, Cham, 2023. Springer International Publishing.

[22] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

[23] Mohammad A. Bsoul, Abdallah Qusef, and Saleh Abu-Soud. Building an optimal dataset for arabic fake news detection. *Procedia Computer Science*, 2022.

[24] Cody Buntain and Jennifer Golbeck. Automatically identifying fake news in popular twitter threads. In *2017 IEEE International Conference on Smart Cloud (SmartCloud)*, pages 208–215, 2017.

[25] Juan Cao, Peng Qi, Qiang Sheng, Tianyun Yang, Junbo Guo, and Jintao Li. Exploring the role of visual content in fake news detection. *Disinformation, Misinformation, and Fake News in Social Media: Emerging Research Challenges and Opportunities*, pages 141–161, 2020.

[26] Yimin Chen, Niall J Conroy, and Victoria L Rubin. Misleading online content: recognizing clickbait as" false news". In *Proceedings of the 2015 ACM on workshop on multimodal deception detection*, pages 15–19, 2015.

[27] Amina Chouigui, Oussama Ben Khiroun, and Bilel Elayeb. Ant corpus: An arabic news text collection for textual classification. *ACS/IEEE International Conference on Computer Systems and Applications*, 2017.

[28] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[29] Mohammed El Manar Righi, Djallel Eddine Boussahel, Djamila Mohdeb, Meriem Laifa, and Messaoud Bendiaf. Rumor stance classification: A case study on the propagation of political rumors on the algerian online social space. In *2022 International Conference on Advanced Aspects of Software Engineering (ICAASE)*, pages 1–6, 2022.

[30] Hanen Himdi, George Weir, Fatmah Assiri, and Hassanin M. Al-Barhamtoshy. Arabic fake news detection based on textual analysis. *Arabian journal for science and engineering*, 2022.

[31] Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin De Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. Parameter-efficient transfer learning for nlp. In *International Conference on Machine Learning*, pages 2790–2799. PMLR, 2019.

[32] Ghaith Jardaneh, Hamed Abdelhaq, Momen Buzz, and Douglas 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)*, pages 596–601, 2019.

[33] Alan S. Kaye, Alan S. Kaye, Salih J. Altoma, and Salih J. Altoma. The problem of diglossia in arabic: A comparative study of classical and iraqi arabic. *Journal of the American Oriental Society*, 1972.

[34] Nassima Kerras, Nassima Kerras, Nassima Kerras, Nassima Kerras, and E Moulay-Lahssan Baya. Standard arabic and algerian languages: A sociolinguistic approach and a grammatical analysis. *null*, 2019.

[35] Ashwaq Khalil, Moath Jarrah, Monther Aldwairi, and Manar Jaradat. Afnd: Arabic fake news dataset for the detection and classification of articles credibility. *Data in Brief*, 2022.

[36] Jude Khouja. Stance prediction and claim verification: An Arabic perspective. In *Proceedings of the Third Workshop on Fact Extraction and VERification (FEVER)*, Seattle, USA, 2020. Association for Computational Linguistics.

[37] David O Klein and Joshua R Wueller. Fake news: A legal perspective. *Australasian Policing*, 10(2), 2018.

[38] Gregory Koch, Richard Zemel, Ruslan Salakhutdinov, et al. Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*, volume 2. Lille, 2015.

[39] Nir Kshetri and Jeffrey M. Voas. The economics of fake news. *IT Professional*, 19:8–12, 2017.

[40] Adam J. Kucharski. Post-truth: Study epidemiology of fake news. *Nature*, 540:525–525, 2016.

[41] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

[42] Mohamed Maamouri, Ann Bies, Tim Buckwalter, and Wigdan Mekki. The penn arabic treebank : Building a large-scale annotated arabic corpus. 2004.

[43] Mohamed Maamouri, Ann Bies, Seth Kulick, Michael Ciul, Nizar Habash, and Ramy Eskander. Developing an Egyptian Arabic treebank: Impact of dialectal morphology on annotation and tool development. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 2348–2354, Reykjavik, Iceland, May 2014. European Language Resources Association (ELRA).

[44] Ahmed Mahlous and Ali Al-Laith. Fake news detection in arabic tweets during the covid-19 pandemic. *International Journal of Advanced Computer Science and Applications*, 12, 07 2021.

[45] Djamila Mohdeb, Meriem Laifa, and Miloud Naidja. An arabic corpus for covid-19 related fake news. *2021 International Conference on Recent Advances in Mathematics and Informatics (ICRAMI)*, 2021.

[46] El Moatez Billah Nagoudi, AbdelRahim Elmadany, Muhammad Abdul-Mageed, and Tariq Alhindi. Machine generation and detection of Arabic manipulated and fake news. In *Proceedings of the Fifth Arabic Natural Language Processing Workshop*, pages 69–84, Barcelona, Spain (Online), December 2020. Association for Computational Linguistics.

[47] Ali Nassif, Ashraf Elnagar, Omar Elgendy, and Yaman Afadar. Arabic fake news detection based on deep contextualized embedding models. 05 2022.

[48] Jonas Pfeiffer, Ivan Vulić, Iryna Gurevych, and Sebastian Ruder. Mad-x: An adapter-based framework for multi-task cross-lingual transfer. *arXiv preprint arXiv:2005.00052*, 2020.

[49] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*, 2019.

[50] Houda Saadane, Houda Saadane, and Nizar Habash. A conventional orthography for algerian arabic. *ANLP@ACL*, 2015.

[51] Hadeel Saadany, Emad Mohamed, Emad Mohamed, Emad Mohamed, Constantin Orasan, and Constantin Orasan. Fake or real? a study of arabic satirical fake news. *arXiv: Computation and Language*, 2020.

[52] Aktham Sawan, Thaer Thaher, and Noor Abu-el rub. Sentiment analysis model for fake news identification in arabic tweets. In *2021 IEEE 15th International Conference on Application of Information and Communication Technologies (AICT)*, pages 1–6, 2021.

[53] Timo Schick and Hinrich Schütze. Exploiting cloze questions for few shot text classification and natural language inference. *arXiv preprint arXiv:2001.07676*, 2020.

[54] Wesam Shishah. Jointbert for detecting arabic fake news. *IEEE Access*, 10:71951–71960, 2022.

[55] SHAYMAA E Sorour and HANAN E Abdelkader. Afnd: Arabic fake news detection with an ensemble deep cnn-lstm model. *J. Theor. Appl. Inf. Technol.*, 100(14):5072–5086, 2022.

[56] Thaer Thaher, Mahmoud Saheb, Hamza Turabieh, and Hamouda Chantar. Intelligent detection of false information in arabic tweets utilizing hybrid harris hawks based feature selection and machine learning models. *Symmetry*, 13:556, 03 2021.

[57] Lewis Tunstall, Nils Reimers, Unso Eun Seo Jo, Luke Bates, Daniel Korat, Moshe Wasserblat, and Oren Pereg. Efficient few-shot learning without prompts. *arXiv preprint arXiv:2209.11055*, 2022.

[58] Ahmet Üstün, Arianna Bisazza, Gosse Bouma, and Gertjan van Noord. Udapter: Language adaptation for truly universal dependency parsing. *arXiv preprint arXiv:2004.14327*, 2020.

[59] Soroush Vosoughi, Deb K. Roy, and Sinan Aral. The spread of true and false news online. *Science*, 359:1146 – 1151, 2018.

[60] Tahseen A. Wotaifi and Ban N. Dhannoon. Improving prediction of arabic fake news using fuzzy logic and modified random forest model. *Karbala international journal of modern science*, 2022.

[61] Ayman Al Zaatari, Rim El Ballouli, Shady ELbassouni, Wassim El-Hajj, Hazem Hajj, Khaled Shaban, Nizar Habash, and Emad Yahya. Arabic corpora for credibility analysis. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 4396–4401, Portorož, Slovenia, May 2016. European Language Resources Association (ELRA).

[62] Xinyi Zhou and Reza Zafarani. A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, 53(5):1–40, 2020.

[63] Arkaitz Zubiaga, Ahmet Aker, Kalina Bontcheva, Maria Liakata, and Rob Procter. Detection and resolution of rumours in social media: A survey. *CoRR*, abs/1704.00656, 2017.