Informatics Bulletin, Faculty of Computers and Artificial Intelligence, Helwan University
Published Online *Vol 6 Issue 2, July 2024*
(https://fcihib.journals.ekb.eg)

Fake News Detection Techniques: A Survey

Mostafa Mahmoud M. Elhawary[1], Doaa S.Elzanfaly[1], Nermin Abdelhakim Othman[1,2]

[1]Information System Dept. Faculty of Computers Artificial Intelligence, Helwan University Cairo, Egypt
[2]Faculty of Informatics and Computer Science, British University in Egypt, Cairo, Egypt
elhawary465@gmail.com , doaa.saad@fci.helwan.edu.eg, drnermin@fci.helwan.edu.eg

***Abstract*—With the spread of smartphones and the use of social communication networks with different groups of people, fake news has been spread to gain some interactions or any other intentions. Unfortunately, people trust social media platforms and believe most in their news even if it is impossible so we can say that any news of any type will find their audience or believers. People's trust and beliefs translated into actions, so fake news canlead to problems that may affect the economy, politics, or panic at the individual level. Organizations with malicious intents exploit this point which can be described as a lake of consciousness to perform their goals which can be beating another competitive organization or destroying a country and displacing innocentpeople. Recently, many studies have shown wide interest in the process of classifying false news. The classification of fake news falls under the classification of texts and is a sub-task of understanding natural language. In this paper, a reference surveyis provided for all the methods and methodologies that have been used by researchers to discover fake news that spread through social network sites. The used datasets, classification techniques, models, and results are discussed.**

I. Introduction

These days, with the spread of the Internet on a large scale and the emergence of social network platforms such as Twitter, Instagram, Facebook, etc., fake news spreads at a very bad speed due to these platforms. The percentage of false news from users on social media platforms is approximately 62 percent [1]. In addition, social media platforms facilitate accessing information and information sharing [2], [3]. Intended fake news which could be written by agencies or persons for financial or political purposes is a critical issue due to its negative impact [4]–[7]. False news also spreads with the aim of ignorance or panic, as happened with theCorona pandemic [8]. The availability of data on social media platforms has garnered a lot of interest among researchers and become a hot spot for fake news sharing. And that through the context of writing false news. The power of social media and its great impact on all of society aspects whether people live, the economy, state security, or state politics become a great fact that imposes itself as a very dangerous weapon that can do more than nuclear weapons can do, from this aspect the attention of fourth generation war issue has arisen again. In the past, fake news was less dangerous because it was spreading very slowly, very restricted to a small region, only a few people were interested in the news and the news couldn't be audience-selective. The growing number of social media platforms facilitate the spread of fake news, and on the contrary, complicate the detection of it. In the few last years, developing countries have been badly affected by fake news. In [9] authors present the limitations of manual detection due to the overwhelming volume of articles globally and emphasize the need for an automated system to assign credibility scores to different publishers and news contexts. Before we study fake news, we should know its types and categories to understand what authors mean in their studies. False information refers to Deceiving information whatever its type like hoaxes, propaganda, rumors, and junk news, so to make our study more organized and useful we can categorize fake news into disinformation which is defined as false information created and shared with malicious aims, misinformation which is spreading false information with no aims, and malinformation despite this type is shared real information, but with malicious intent [10]. Furthermore, we can define more minor types of false news such as fake news which is defined as Feigned articles that could be potentially or purposely deceptive to the audience, a hoax which is a trick that spreads containing false or inaccurate news, propaganda which is defined as an organized campaign to influence the audience to make them think about or do a particular thing, a satire which is created in form of jokes to amuse the audience. It may be harmful if it contains deceptive news, and rumors which can be defined as a subclass of propaganda. Rumors can be shared from one person to another without knowing the source of them, and click-bait which is attractive news used by the low-level press to make traffic to their sites to gain revenue [10].

There are two approaches researchers have used to detect fake news. The first approach is by extracting the features in the text and then introducing them to machine learning (ML) techniques. In [11] authors investigate five ML classifiers: random forest (RF) [12], Naïve Bayes (NB) [13], Support Vector Machine (SVM) [14], Logistic Regression (LR) [15], and Decision Tree (DT) [16]. In [17] authors perform feature extraction on the Kurdish dataset which was extracted from Facebook using a Facebook-scraper tool using Term Frequency-Inverse Document Frequency (TF-IDF). In [18] authors investigate three ML classifiers: Multinomial Naïve Bayes (MNB) [19], DT, and SVM on a dataset of Arabic content consisting of YouTube comments. They convert the dataset texts into features of n-gram size of words, then they apply TF-IDF to the extracted features. In [20] authors use SVM for fake news classification with dataset [21] which contains seven classes corresponding to each news to describe it. They use class two and class six with the method described in [22]. In [8] authors investigate five ML classifiers: NB, LR, Multilayer Perceptron (MLP) [23], RF, eXtreme Gradient Boosting Model (XGB) [24] with TF-IDF and word count features, and apply them to an Arabic dataset scraped from Twitter.

The second approach relies on the use of deep learning to identify fake news. In [24] authors use the multi-model based on Neural Networks. The model detects fake news with two inputs the image and the text of the news. In [25] authors try three deep neural network variants to detect fake news by the input text, which is Long-Short Term Memory (LSTM) [26], LSTM with dropout regularization, and LSTM with one-dimension convolutional neural networks (1D-CNN) [27]. In [28] authors use the Bidirectional LSTM (BI-LSTM) [29] and autoencoder for fake news detection by input text. In [30] authors use the Arabic Bidirectional Encoder Representations from Transformers (AraBERT) [31] with a dataset in [20]. The AraBERT model is a pre-training model built based on Transformer [32] with Arabic content. In [33] authors use the Bidirectional Encoder Representations from Transformers (BERT) [34] for fake news detection by input text. The BERT model is a pre-training model built based on Transformer with English content.

The dataset is a very important factor that affects the detection model efficiency, so it should be processed and purified as much as possible. Researchers in any language but English suffer from lacking benchmark datasets, so many researchers have created their datasets and then trained their models on them. There are popular English datasets like LIAR [35], Twitter15 [36], and Weibo [37]. There are also Arabic datasets like Covid19Fakes [38], AraNews [39], and ANS[40]. Datasets vary depending on their format, classes, source, collecting time, size, and subject. Twitter15 and LIAR are very popular English datasets that are publicly available [41]. The AraNews dataset is a very large one containing various topics collected from 5o newspapers from 15 Arabic countries, the United States of America (USA), and the United Kingdom (UK), it contains more than 5 million news articles. The ANS dataset is an Arabic dataset containing 3072 real samples versus 1475 fake samples with a total of 4547 samples [40]. The Covid19Fakes consisted of an English dataset and an Arabic dataset in CSV format, the author provides the IDs of each tweet and its label. Covid19Fakes contains more than 200000 English tweets and more than 200000 Arabic tweets. The purpose of this paper is to provide a reference survey of researchers who have provided methods and methodologiesfor identifying fake news on social media platforms.

This paper is divided into five sections: Section I Introduction, Section II Fake News Detection Approaches, Section III Discussion, Section IV Challenges and Research Future Directions, and Section V Conclusion.

II. Fake news detection approaches

Many automatic fake news detection approaches have been proposed in the last few years due to their harmful impact. In [42] authors present a comparison among methods based on different approaches like traditional ML and DL models. We can categorize these approaches into two main sections. The first section includes approaches based on machine learning which suffer from feature extraction steps, so many researchers turned to deep learning approaches which are the second approach. Deep neural networks simplify feature extractionby automatically learning hierarchical and abstract representations directly from the raw data reducing the need for manual intervention and domain specific knowledge.

A. Classification of fake news using ML

Machine learning has a great history in the natural language processing field, but it needs a necessary step called the feature extraction stage, this stage consumes computer resources and time. Researchers need to minimize the algorithm complexity to save time and resources and maximize performance. Before the feature extraction stage, there is another important step called data pre-processing. Generally, data pre-processing steps are almost the same in all languages, but we should consider the language families. We can summarize common pre-processing steps regardless of the language as follows:

1) Remove HTML tags: news articles often contain HTML tags for formatting. We need to remove these tags and extract just the text content. This can be done using the BeautifulSoup library in Python.
2) Remove punctuation and numbers: we remove punctuation marks and numbers from the text since they do not contribute much to the meaning. We can use a regular expression to remove these.
3) Remove stop words like “the”, “a”, and “an” whichare very common but do not provide much useful information.
4) Lemmatize words like “fishes”, “went”, and “better”have the same lemma as “fish”, “go”, and “good”.
5) Convert to lowercase: we convert all characters to lowercase so that the model does not treat the same words with different cases differently.
6) Check for null values and duplicates: we should also check for any null values or duplicates in the data and handle them appropriately.

Preprocessing Arabic text for fake news detection purposes has some additional challenges compared to English. Some of these challenges are:

1) Removing diacritics: Arabic text contains diacritics like fathas, dhammas, and kasra to represent short vowels. We remove these diacritics since they do not contribute to the meaning.
2) Normalizing different forms: the Arabic script has different forms for the same letter depending on its position in the word. We need to normalize the letters to a single form. This can be done using libraries like ArabicNormalize.
3) Removing stop words: Like in English, we remove common stop words like: لل, الذى, في.
4) Stemming: Arabic words are derived from a set of roots by applying different prefixes and suffixes. Stemming aims to remove these affixes and normalize the word to its root form يشربون, تفعل turned to شرب, فعل.
5) Handling encoding: there are different encodings for Arabic scripts like CP1256, ISO 8859-6, UTF-8, etc. We need to ensure our text is in a consistent encoding (typically UTF-8) to avoid errors. After dataset pre-processing and feature extraction, features are passed with their labels to the classifier to know if it is fake or not. Figure 1 presents the fake news detection phases.

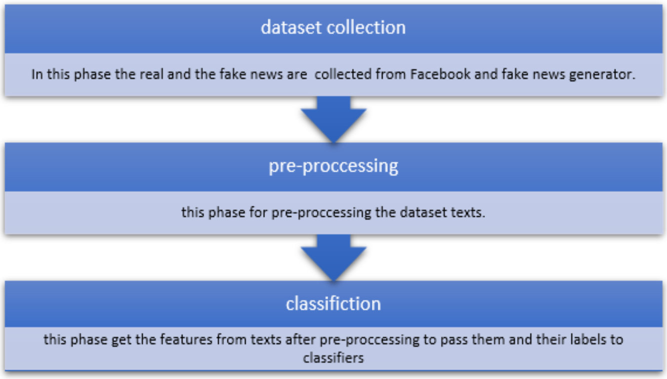
dataset collection
In this phase the real and the fake news are collected from Facebook and fake news generator.
pre-proccessing
this phase for pre-processing the dataset texts.
classifiction
this phase get the features from texts after pre-proccessing to pass them and their labels to classifiers

Fig. 1: fake news detection phases

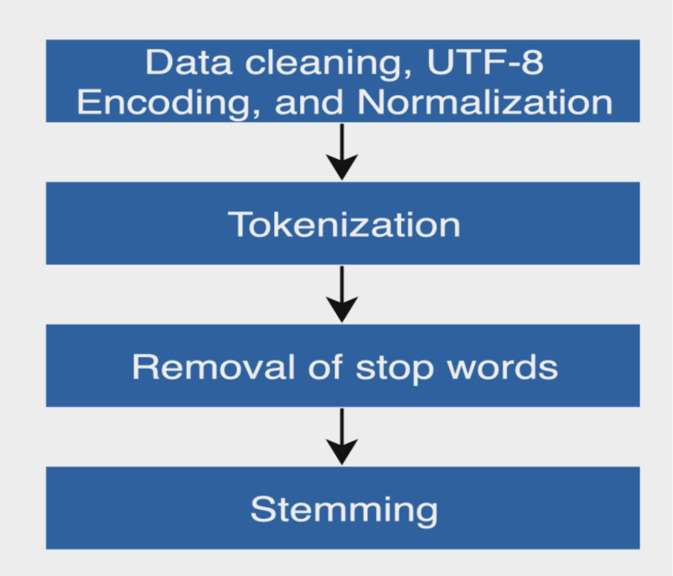

Fig. 2: The Sequence of Pre-processing Phase

In this section, we will show researchers' efforts in machine learning across different languages, including both Arabic and English.

1) Kurdish language: in [11] authors use ML Techniques for detecting fake news in the Kurdish language which includes three phases. The first phase is the dataset collection, in this phase authors use two sets, the first set is a collected real news and fake news in the Kurdish language using the Facebook tool. The second set is made by modifying the real news in the first set to generate the fake news. The second phase is the dataset pre-processing. The pre-processingcontains four consecutive units. The first unit is encoding and normalization for encoding the text to UTF 8 and removing the characters such as: *, @, %, &..., URL links, the non- Kurdish words, emojis, and extra spaces. The second unitis tokenization for splitting the text in natural language into vectors containing words and tokens in a special language. The tokenization unit in [11] needs ontology content words in the Kurdish language. The third unit is for removing the stop words in the sentence. Any language has to stop words suchas "is, are, and, or, etc." in English. The Kurdish languagehas 240 stop words [43]. The last unit is stemming words in the Kurdish language. The stemming is a process for getting the root of the word. Figure 2 presents the sequence of the pre-processing phase. The last phase is the classifier for the classification of fake news and real news, but there is feature extraction before the classification. In [11] authors use the TF-IDF to extract features of sentences, then pass the label and feature to the classifier. In [11] authors investigate five ML classifiers: RF, NB, SVM, LR, and DT. The first set achieves 88.71% accuracy with SVM, and the second set achieves 83.26% accuracy with RF.

2) Arabic language: In [18] authors use ML to detect fake news about celebrities' deaths in YouTube comments. There are four phases to detect fake news. The first phase is the data collection from comments on YouTube about three celebrities in three sets. Distribute the news in the three sets into fake and non-fake news about a celebrity's death. The second phase is data cleaning which is the same phase as in [11]. The third phase is feature extraction using TF-IDF on sentences to convert words to vectors. The last phase is passing the features with labels to investigate by three ML classifiers: SVM, DT, and MNB. Figure 3 presents the pipeline of detecting fake news processes. The performance of detection of fake news achieves 95.35% accuracy with SVM for the first celebrity fake news. The second celebrity fake news achieves 95.56% with DT. The third celebrity fake news achieves 93.68% with SVM. In [20] authors use the dataset in [21], which has seven types of classes , type 1 contains a verifiable factual claim, type 2 contains false information, type 3 is attractive to the public, type 4 is possible damage to a person, an organization, or the community, type 5 needs reviewing by a fact-checker, type 6 harms society, type 7 needs policymakers' attention. Type 2 and type 6 are related to fake news. In [20] authors use the SVM for the classification of the six types of classes. Type 2 achieves 44.3% accuracy with Arabic content and 40% accuracy with English content. Type 6 achieves 69.7% accuracy with Arabic content and 50% accuracy with English content.

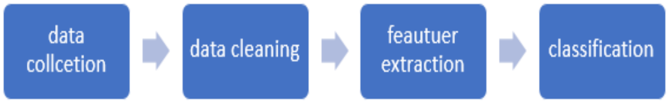

Fig. 3: The Pipeline of Fake News Classification Process

In [8] authors use ML to detect fake news about Covid-19 on Twitter. There are three steps for detecting fake news in Covid-19. The first step is data collection from Twitter by Hashtags such as الحجر الصحي #, كورونا #, etc. which contain real and fake news about COVID-19, then they manually annotated 2,500 tweets into fake or genuine classes. After annotation, there are 1,537 tweets (835 fake and 702 genuine). In the first step there is a second phase which is tweets automatically annotated by ML from the Manual Annotated dataset, the tweets automatically Annotated were 34,529 tweets (19,582 fake and 19,582 genuine). The feature extraction step investigates four types of features: Count Vector, Word-Level TF-IDF, N-gram-Level TF-IDF, and Character-Level TF-IDF. The last step is classification by ML, there are six ML investigation algorithms NB, LR, SVM, MLP, RF, and XGB. The best f1-score is 87.8% with TF-IDF (n-gram-level) feature and LR classifier. Table 1 presents each of the Arabic fake news detection dataset results on YouTube comments and Twitter with used features and classifiers.

3) English, Spanish, and Portuguese language: in [44] authors use ML to detect fake news in three languages English, Spanish, and Portuguese. They applied three steps pre-processing, feature extraction, and classification for predicting fake news. The first step is pre-processing which has three phases cleaning, filtering, and noise removal. In the cleaning phase, the text is converted to UTF-8 then removing non-textual characters and special characters. The filtering phase filters the small text to avoid too short news. The noise removal phase removes extra whitespaces and removes the text that is not related to the news content from the original text. In the second step three categories of features are used: complexity, stylistic, and psychological as in [45], these features focus on capturing high-level structures. The complexity features capture the overall intricacy of the news at the word and sentence level. The stylistic feature uses NLP to extract grammatical information from each document or sentence, also, the stylistic feature is understood as a syntax and text style using NLP techniques. The psychological features use sentiment polarity evaluation [46] for measuring the positivity or negativity of a text. The third step uses SVM, k-NN, RF, and XGB for classification. There are three classes: Fake, Legitimate, and satire in the three datasets of fake news. The three datasets are English fake news [45], Portuguese Fake news [47], and Spanish Fake news [48]. In [44] authors archive between 71% and 91% accuracy score in the English dataset, 89% accuracy score in the Portuguese dataset, and 77% accuracy score in the Spanish dataset.

B. Classification of fake news using DL

Deep learning has dominated the research community due to its better performance compared to traditional machine learning in many domains. Deep learning is characterized by its different representations according to the number of levels and the way the neurons are connected whether in the same level or between one level and another.

Here we will demonstrate how researchers have applied deep learning techniques to various languages like English and Arabic.

1) English language: in [25] authors present a multi-model based on CNN for feature extraction and a neural network for classification of the fake news and event discriminator which has two inputs for detecting the information of the event. The first input is the text of the event which is converted into a vector using the word embedding then they use 1D-CNN to extract the features from word embedding. The second input is the image of the event which the authors extract the features from it using the VGG-19 model [49]. The VGG-19 is a pre-trained model for feature extraction and detection from images. After extracting the features, the features from the text and image are concatenated into the feature model. Then use the feature model for news classification to decide whether the event is fake news or not. The event discriminator is for detecting what the event is. News classification and event discriminators use the neural network for prediction. In [25] authors present a model called Event Adversarial Neural Networks (EANN). Figure 4 presents the methodology of EANN. The dataset in [25] is a collection of two datasets, the first dataset collection is from Twitter [50] for detecting fake content on Twitter and the second dataset is called Weibo Dataset [51] used for fake news detection. Table 2 presents the number of samples and classes in each dataset Twitter and Weibo. The model achieves 71.5% accuracy with the Twitter dataset and 82.7% with Weibo. In [26] authors try three different deep learning models trained by a dataset from Twitter content presented in [52]. The dataset contains 5800 tweets about five rumor stories. Figure 5 shows the three deep-learning models applied to the dataset. The first model includes five steps, the first step is word embedding to convert the word and token to vector. The second step uses the CNN model to reduce the dimensionality of embedding word vectors and avoid over-fitting of the training data. The third step uses the output feature from CNN to train LSTM. The fourth step is a dense layer (full connection) to connect the output vectors of LSTM to the signal layer. The last step is classification to predict the label based on a neural network.The second model contains four steps: word embedding, the LSTM with dropout, dense layer (full connection), and classification. The last model contains word embedding, LSTM, dense layer (full connection), and classification. The last model achieves 82.29% accuracy after cross-validation training. In [28] authors present a model for fake news detection based on Bi-LSTM and Autoencoder trained using a dataset called the fake news challenge FNC-1[53]. The model depends on two

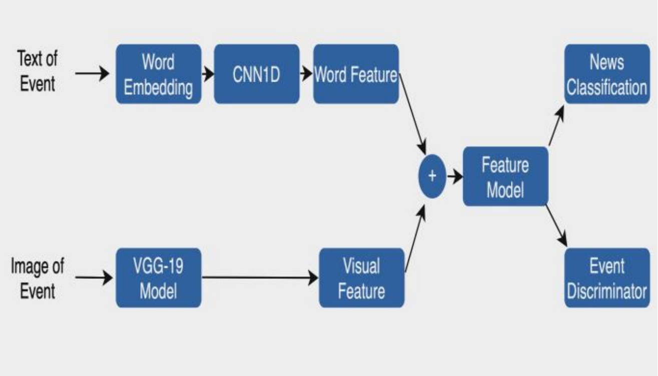

Fig. 4: The Methodology of EANN for Classification Fake News and Event Discriminator

TABLE I: The best results for Arabic fake news datasets, with types of features and classifiers

Study	Dataset	Feature	Approach	Classifier	Best accuracy or f1
M. Alkhair et al. [18]	YouTube comment [18]	TF-IDF	Machine learning	SVM, DT, and MNB	95.35% for 1st celebrity, 95.56% for 2nd celebrity, and 93.68% for 3rd celebrity
Firoj Alam et al. [21]	Twitter dataset [21]	TF-IDF	Machine learning	SVM	44.3 % for type 2, and 69.7% for type 6
Ahmed Redh et al. [8]	Covid-19 Twitter dataset [8]	Count Vector, Word-Level TF-IDF, N-gram-Level TF-IDF, and Character-Level TF-IDF	Machine learning	NB, LR, SVM, MLP, RF, and XGB	f1-scour 87.8%

TABLE II: The number of samples in Twitter and Weibo datasets [25]

Dataset	Twitter	Weibo
Number of fake News	7898	4749
Number of real News	6026	4779
Number of images	514	9528

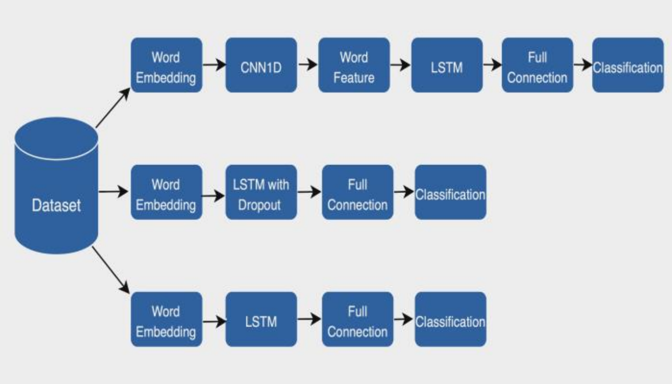

Fig. 5: Present the Three Models of Deep Learning

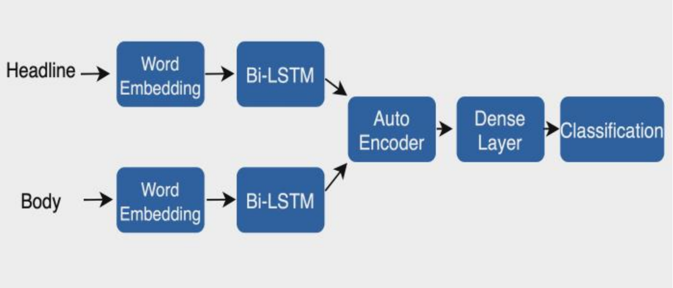

Fig. 6: The methodology of fake news detection for FNC dataset

inputs of news the headline and the body of news. For each, the inputs need to be converted to vectors by word embedding, then the Bi-LSTM for feature extraction from them. The Bi-LSTM is a double LSTM in two directions, then use the Autoencoder for connecting the two outputs from the Bi-LSTM. The output from the autoencoder passes to the dense layer (full connection) for normalization and linear vector, then classifies the news using the classification stage into four classes Agree, Disagree, Discuss, and Unrelated. Figure 6 presents the methodology of the model. The model achieves 94% accuracy, 93.725% Precision, 93.88%, recall, and 93.88% F1 Score.

In [31] authors use the BERT which is a pre-trained model based on transformer training on the BookCorups dataset which is made of over 800 million words to extract features from input sentences. The BERT has a bidirectional trans- former encoder in each layer. The transformer encoder unit has three components: multi-head attention, add and normalize, and feed-forward. Figure 7 shows the BERT architecture.

In [33] authors propose two models based on BERT for fake news classification using LAIR [35] and LAIR PLUS [54] datasets. The first model for fake news classification using the LAIR dataset has two input branches because the LAIR dataset has a news statement and metadata for each news. The first model uses the BERT for two branches of the dataset then shares the weights between them and concatenates the output of BERT then uses the dropout to avoid over-fitting. Finally, they use the full connection and soft-max function for output classification. Figure 8 presents the architecture of the first model for the LAIR dataset.

The second model for fake news classification using the LAIR PLUS dataset has three input branches because the LAIR PLUS dataset has a news statement, metadata, and human justification for each piece of news. The second model uses the BERT for the three branches of the dataset then shares the weights between them and concatenates the output of BERTs, then uses the Dropout to avoid the over-fitting. Finally, the authors use the full connection and soft-max function for output classification. Figure 9 presents the architecture of the second model with the LAIR PLUS dataset. There are two mods in LAIR and LAIR PLUS, the first mod has a binary label ‘true’ and false. The second mod has six labels ‘true’, ‘half true’, ‘mostly true’, ’mostly false’, ‘false’, and ‘pants- fire’. The first model achieves 72% accuracy with binary labels in classification and 35.2% with six labels in classification. The second model achieves 74% accuracy with binary labels in classification and 37.1% with six labels in classification. In [55] authors achieve an F1-score of 66.67% using the BERT algorithm, but they have used a very small dataset. In [56] authors implemented a model called (FakeBERT) which consisted mainly of BERT with 1D-CNN followed by a max-pooling layer. FakeBERT achieved an accuracy of 98.9% with 10 epochs. In [57] authors use LSTM with a glove word embedding to implement their model, then they built the dataset

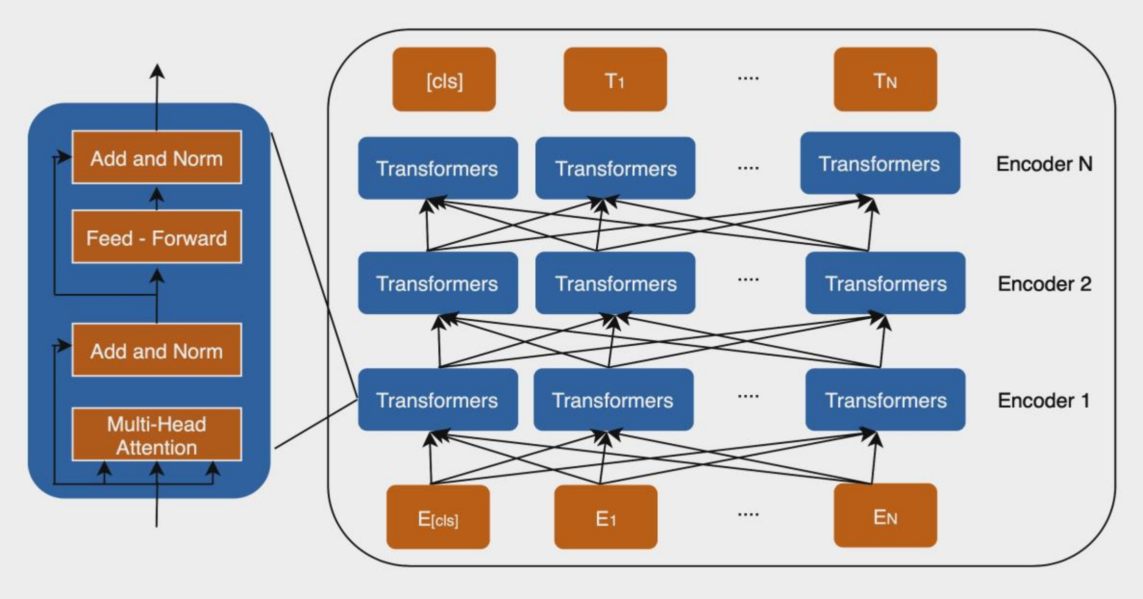

Fig. 7: BERT Architecture

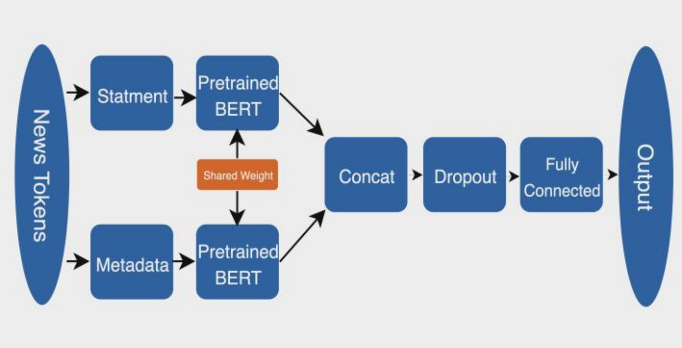

Fig. 8: The first Model for Fake News classification from LAIR Dataset [33]

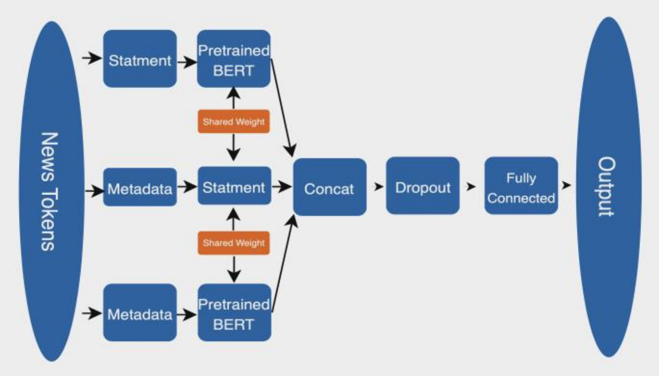

Fig. 9: The Second Model for Classification Fake News from LAIR PLUS Dataset [33]

from two sources one of them is from Kaggle.com, and the other one is the gloveTwitter data which is used in embedding. They achieve an accuracy of 99.88% with 10 epochs. Table 3 presents some English datasets for fake news detection using the DL techniques and the results.

2) Arabic language: in [30] authors use the AraBERT for fake news detection from Twitter with Arabic dataset content. Figure 10 presents the methodology of the model. The AraBERT is a pre-trained model with Arabic content that has the special tokens: [SEP] for segment separation and [CLS] for classification that is used as the first input token for any classifier. Authors use the output of [CLS] for classification by connecting with a feed-forward neural network and then use the sigmoid function to normalize the output between 0 and 1. In [30] the model was trained by the dataset used in [21]. The dataset in [21] has seven types of information, but only types 2 and 6 are related to fake news. The model achieves 67.7% accuracy, 78.7% precision, 67.7% recall, and 66.4% F1 score for type 2. And achieves 90% accuracy, 90% precision, 91% recall, and 90% F1 score for type 6. In [58] authors use the CNN and LSTM for fake news detection from Twitter with the ANS dataset [40]. Figure 11 presents the methodology of the model. The first step embedding the text to convert words to vectors. The second step passes embedding words into CNN for convolution and pooling based on one dimension of the embedding words. The CNN model works as a redundancy of the feature of embedding words. The third step passes the output of the CNN model into the LSTM model. The LSTM model for learning from the sequence of words. The last step uses the dense (full connection) layer for labeling the output of LSTM. The ANS dataset contains 4547 samples where 1475 samples are real news and 3152 samples are fake news. This model achieves 67% accuracy, 49% precision, 57% recall, and 52% F1-score. In [59] upgrade the Arabert and mBERT [60] by fine-tuning with 1.5 million Arabic sentences about covid 19 [61] before being used with a labeled dataset. This model is called after fine-tuning with COVID-19 Arabic

TABLE III: The Best Results for English Fake News Datasets Using DL Techniques

Study	Dataset	approach	Techniques	Best accuracy
Jin, Z., Cao, J. et al. [51]	Twitter [50] Weibo Dataset [51]	Deep learning	VGG-19 and Event Adversarial Neural Networks	71.5 % 82.7 %
Zubiaga et al. [52]	Twitter dataset [52]	Deep learning	LSTM	82.29 %
Slovikovskaya et al. [53]	FNC-1 dataset [53]	Deep learning	Bi-LSTM and Auto encoder	94%
Wang, W. Y. et al [35]	LAIR [35]	Deep learning	BERT	72%
Alhindi, T et al. [54]	LAIR PLUS [54]	Deep learning	BERT	74%

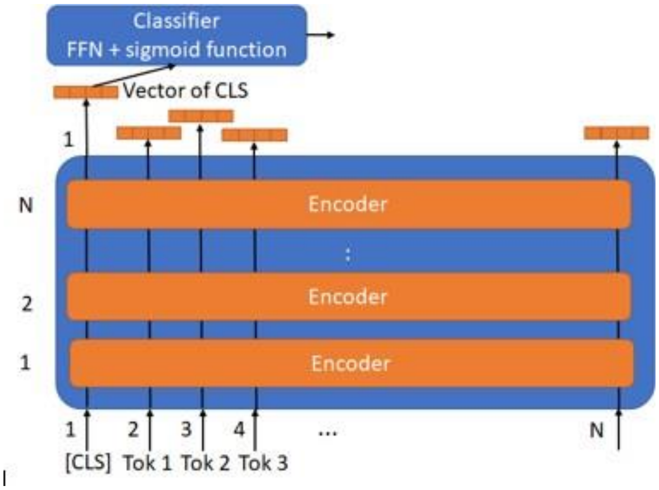

Fig. 10: The Methodology of Fake News Detection for Arabic Content on Twitter

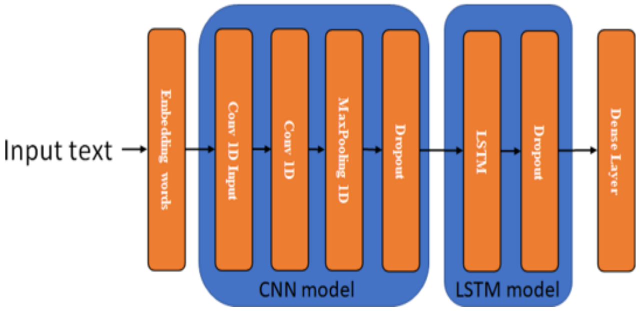

Fig. 11: The Methodology of Fake News Detection based on CNN-LSTM using ANS Dataset

sentences by AraBert cov19 from AraBert and mBert cov19 from mBert. This model works as the model in [30] but is fine-tuned with a different dataset. The used one consists of 10828 Arabic tweets classified with 10 distinct labels. The labels are about hate, cures, giving advice, morals, judgments or news, dialect, lambaste and negative speech, factual, need fact-checking, and contain false information. This study achieves an accuracy of 95.78% with the label "contains fake information" using the model "AraBert cov19" and 94.91% when using the model "mBert cov19" for the same label. Table 4 presents each ofthe datasets for Arabic fake news detection results using the DL techniques.

TABLE IV: The Best Results for Arabic Fake News Datasets using DL Techniques

Dataset	Techniques	Best accuracy
dataset [21]	AraBERT	90%
ANS dataset [40]	CNN-LSTM	67%
AraCOVID19-MFH [59]	AraBERT	94.17 %
	mBERT	93.53 %
	AraBERT Cov19	95.78 %
	mBERT Cov19	94.91 %

III. Discussion

From the above study, using deep learning shows better results than machine learning algorithms, which is clear in [20] and [30] for the same dataset [21], where the accuracy is 69% in type 6 and 40% in type 2 [20], while the accuracy of the results in deep learning using AraBERT was 67% for type2 and 90% for type 6 [30].

It is not possible to judge the remainder of the deep learning mechanisms that have the same target classifying the fake news, because they use different datasets from each other, whether it is similar in their structure or the number of samples they contain. However, the comparative study [62] applied to different datasets to classify texts. The results of [62] show the superiority of deep learning over machine learning, in addition to the superiority of transformer-based models such as BERT, XLNet [63], RoBERTa [64], and MT-DNN [65] over deep learning models in text classification. In [32] authors present the complexity of deep learning algorithms for NLP and the results show the transformer is superior in training time over other NLP deep learning algorithms such as LSTM, Bi-LSTM, and CNN.

There are limitations in ML as the ML needs a feature extraction step, and there are many algorithms for feature extraction for example, in NLP the feature extraction algorithms used are TF-IDF, n-grams, word counts, etc. but in DL the features are automatically picked by the neural network. Another limitation of ML is about bad performance with complex problems. The ML is not a good choice with large datasets whereas the DL is very good with large datasets, also the ML can't handle the multi-complex inputs in a dataset. Table 5 shows a summary of researchers' studies to facilitate understanding them and compare their results.

IV. Challenges and research future directions

We can note that there are a number of joint challenges among researchers in this field especially those who work

TABLE V: Discussion Table

Study	Dataset	Feature	Approach	Language	Techniques	Best accuracy
Jin, Z., Cao, J. et al. [51]	Twitter [50] Weibo dataset [51]	-	DL	English	VGG-19 and Event Adversarial Neural Networks	71.5 % 82.7 %
Zubiaga et al. [52]	Twitter dataset [52]	-	DL	English	LSTM	82.29 %
Slovikovskaya et al. [53]	FNC-1 dataset [53]	-	DL	English	Bi-LSTM and Auto encoder	94%
Wang, W. Y. et al [35]	LAIR [35]	-	DL	English	BERT	72%
Alhindi, T et al. [54]	LAIR PLUS [54]	-	DL	English	BERT	74%
Abonizio et al. [44]	English fake news [45]	the features focus on capturing high-level structures: Complexity, Stylistic, and Psychological	ML	English	KNN SVM RF XGB	75% 79% 79.9% 80.3%
M. Alkhair et al. [18]	YouTube comment [18]	TF-IDF	ML	Arabic	SVM, DT, and MNB	95.35% for 1st celebrity, 95.56% for 2nd celebrity, and 93.68% for 3rd celebrity
Firoj Alam et al. [21]	Twitter dataset [21]	TF-IDF	ML	Arabic	SVM	44.3 % for type 2, and 69.7% for type 6
Ahmed Redh et al. [8]	Covid-19 Twitter dataset [8]	Count Vector, Word-Level TF-IDF, N-gram-Level TF-IDF, and Character-Level TF-IDF	ML	Arabic	NB, LR, SVM, MLP, RF, and XGB	f1-score 87.8%
Hussein et al. [30]	Dataset [21]	-	DL	Arabic	AraBERT	90%
Sorour et al. [58]	ANS dataset [40]	-	DL	Arabic	CNN-LSTM	67%
Pal, A.P et al. [55]	Dataset [55]	-	DL	English	BERT	F1-score 66.67%
Rohit K. et al. [56]	Real-world fake news dataset [56]	-	DL	English	BERT +1D-CNN	98.9%
Tavishee et al. [57]	Glove twitter dataset [57]	-	DL	English	LSTM	99.88%
Ameur et al. [60]	AraCOVID19-MFH [60]	-	DL	Arabic	AraBERT mBERT AraBERT Cov19 mBERT Cov19	94.17 % 93.53 % 95.78 % 94.91 %

on the Arabic language. The biggest challenge is the Arabic dataset, as most of the datasets are hard to scrap from social sites. Most researchers in the Arabic branch present only the tweets/posts IDs and their class (ex: real or fake), any other researcher needs to scrap the dataset using the IDs. The scrapped dataset can't be the same every time because of the removed tweets by the user. After scrapping, the dataset must be cleaned and purified. The mentioned tasks are very exhaustive and need permissions which makes it a very hard task, so we can say there is an urgent need for an Arabic benchmark dataset that must be publicly available for upcoming research. Another problem in the Arabic dataset is the small size of most of them [41]. Another important issue is how to detect malicious news quickly before spreading [10]. Using ML to extract features from datasets gives big-size vectors that must be minimized to facilitate using it in fake news detection[10]. Fake news detection is a very wide field and still needs more research to cover it. In the future, we can compare the results of different techniques using different datasets to determine the best techniques, then try different evaluation metrics [10]. Ensemble techniques may play an important role in enhancing results [66]. Fake news detection will expand to include DeepFake [67].

CONCLUSION

This research presents the techniques used in categorizing news between true and fake. These techniques are based on one of the following two main approaches: the traditional approach (feature extraction and ML) and the approach based on deep learning. The traditional methodology relies on techniques for extracting features from text such as (TF-IDF, n-gram, count vectors, etc.), then classifying the features using machine learning techniques. Recent research is based on using deep learning to classify news, such as using CNN, RNN, and LSTM. However, the pre-trained models based on Transformer, such as BERT and others have showed better results. In addition, the Transformer is faster in training than other models, and this is an important advantage in the complexity of the algorithms. The criteria used to measure results were accuracy, precision, F1 scores, and recall. We recommend using the Transformer or pre-trained models based on the

Transformer for the fake news classification process.

References

[1] H. Allcott and M. Gentzkow, “Social media and fake news in the 2016 election,” *Journal of economic perspectives*, vol. 31, no. 2, 2017.

[2] M. Del Vicario, A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, H. E. Stanley, and W. Quattrociocchi, “The spreading of misinformation online,” *Proceedings of the national academy of Sciences*, vol. 113, no. 3, pp. 554–559, 2016.

[3] S. Kumar and N. Shah, “False information on web and social media: A survey,” *arXiv preprint arXiv:1804.08559*, 2018.

[4] S. Ghosh and C. Shah, “Towards automatic fake news classification,” *Proceedings of the Association for Information Science and Technology*, vol. 55, no. 1, pp. 805–807, 2018.

[5] N. Ruchansky, S. Seo, and Y. Liu, “Csi: A hybrid deep model for fake news detection,” in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2017, pp. 797–806.

[6] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, “Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media,” *Big data*, vol. 8, no. 3, pp. 171–188, 2020.

[7] X. Zhou and R. Zafarani, “A survey of fake news: Fundamental the- ories, detection methods, and opportunities,” *ACM Computing Surveys (CSUR)*, vol. 53, no. 5, pp. 1–40, 2020.

[8] A. R. Mahlous and A. Al-Laith, “Fake news detection in arabic tweets during the covid-19 pandemic,” *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 778–788, 2021.

[9] Z. Khanam, B. Alwasel, H. Sirafi, and M. Rashid, “Fake news detection using machine learning approaches,” in *IOP conference series: materials science and engineering*, vol. 1099, no. 1. IOP Publishing, 2021, p. 012040.

[10] M. K. Elhadad, K. F. Li, and F. Gebali, “Fake news detection on social media: a systematic survey,” in *2019 IEEE Pacific Rim conference on communications, computers and signal processing (PACRIM)*. IEEE, 2019, pp. 1–8.

[11] R. Azad, B. Mohammed, R. Mahmud, L. Zrar, and S. Sdiqa, “Fake news detection in low-resourced languages “kurdish language” using machine learning algorithms,” *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 6, pp. 4219–4225, 2021.

[12] T. K. Ho, “The random subspace method for constructing decision forests,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 20, no. 8, pp. 832–844, 1998.

[13] H. Zhang, “The optimality of naive bayes,” *Aa*, vol. 1, no. 2, p. 3, 2004.

[14] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine learning*, vol. 20, pp. 273–297, 1995.

[15] J. Tolles and W. J. Meurer, “Logistic regression: relating patient characteristics to outcomes,” *Jama*, vol. 316, no. 5, pp. 533–534, 2016.

[16] B. Kamiński, M. Jakubczyk, and P. Szufel, “A framework for sensitivity analysis of decision trees,” *Central European journal of operations research*, vol. 26, pp. 135–159, 2018.

[17] J. Beel, B. Gipp, S. Langer, and C. Breitinger, “Paper recommender systems: a literature survey,” *International Journal on Digital Libraries*, vol. 17, pp. 305–338, 2016.

[18] M. Alkhair, K. Meftouh, K. Sma¨ıli, and N. Othman, “An arabic corpus of fake news: Collection, analysis and classification,” in *Arabic Language Processing: From Theory to Practice: 7th International Conference, ICALP 2019, Nancy, France, October 16–17, 2019, Proceedings 7*. Springer, 2019, pp. 292–302.

[19] L. Jiang, S. Wang, C. Li, and L. Zhang, “Structure extended multinomial naive bayes,” *Information Sciences*, vol. 329, pp. 346–356, 2016.

[20] P. Nakov, F. Alam, S. Shaar, G. D. S. Martino, and Y. Zhang, “A second pandemic? analysis of fake news about covid-19 vaccines in qatar,” *arXiv preprint arXiv:2109.11372*, 2021.

[21] F. Alam, H. Sajjad, M. Imran, and F. Ofli, “Crisisbench: Benchmarking crisis-related social media datasets for humanitarian information processing,” in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 15, 2021, pp. 923–932.

[22] F. Alam, F. Dalvi, S. Shaar, N. Durrani, H. Mubarak, A. Nikolov, G. Da San Martino, A. Abdelali, H. Sajjad, K. Darwish *et al.*, “Fighting the covid-19 infodemic in social media: a holistic perspective and a call to arms,” in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 15, 2021, pp. 913–922.

[23] F. Rosenblatt *et al.*, *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms*. Spartan books Washington, DC, 1962, vol. 55.

[24] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,”in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[25] Y. Wang, F. Ma, Z. Jin, Y. Yuan, G. Xun, K. Jha, L. Su, and J. Gao, “Eann: Event adversarial neural networks for multi-modal fake news detection,” in *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining*, 2018, pp. 849–857.

[26] O. Ajao, D. Bhowmik, and S. Zargari, “Fake news identification on twitter with hybrid cnn and rnn models,” in *Proceedings of the 9th international conference on social media and society*, 2018, pp. 226– 230.

[27] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[28] A. Karpathy and L. Fei-Fei, “Deep visual-semantic alignments for generating image descriptions,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3128–3137.

[29] S. M. Padnekar, G. S. Kumar, and P. Deepak, “Bilstm-autoencoder architecture for stance prediction,” in *2020 International Conference on Data Science and Engineering (ICDSE)*. IEEE, 2020, pp. 1–5.

[30] A. Hussein, N. Ghneim, and A. Joukhadar, “Damascusteam atnlp4if2021: Fighting the arabic covid-19 infodemic on twitter usingarabert,” in *Proceedings of the Fourth Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda*, 2021, pp. 93– 98.

[31] W. Antoun, F. Baly, and H. Hajj, “Arabert: Transformer-based model for arabic language understanding,” *arXiv preprint arXiv:2003.00104*, 2020.

[32] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, . Kaiser, and I. Polosukhin, “Attention is all you need,” *Advances in neural information processing systems*, vol. 30, 2017.

[33] D. Mehta, A. Dwivedi, A. Patra, and M. Anand Kumar, “A transformer-based architecture for fake news classification,” *Social network analysis and mining*, vol. 11, pp. 1–12, 2021.

[34] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” *arXiv preprint arXiv:1810.04805*, 2018.

[35] W. Y. Wang, “” liar, liar pants on fire”: A new benchmark dataset for fake news detection,” *arXiv preprint arXiv:1705.00648*, 2017.

[36] J. Ma, W. Gao, and K.-F. Wong, “Detect rumors in microblog posts using propagation structure via kernel learning.” Association for Computational Linguistics, 2017.

[37] J. Ma, W. Gao, P. Mitra, S. Kwon, B. J. Jansen, K.-F. Wong, and M. Cha, “Detecting rumors from microblogs with recurrent neural networks,” 2016.

[38] M. K. Elhadad, K. F. Li, and F. Gebali, “Covid-19-fakes: A twitter (arabic/english) dataset for detecting misleading information on covid-19,” in *Advances in Intelligent Networking and Collaborative Systems: The 12th International Conference on Intelligent Networking and Collaborative Systems (INCoS-2020) 12*. Springer, 2021, pp. 256–268.

[39] E. M. B. Nagoudi, A. Elmadany, M. Abdul-Mageed, T. Alhindi, and H. Cavusoglu, “Machine generation and detection of arabic manipulated and fake news,” *arXiv preprint arXiv:2011.03092*, 2020.

[40] J. Khouja, “Stance prediction and claim verification: An arabic perspective,” *arXiv preprint arXiv:2005.10410*, 2020.

[41] M. F. Mridha, A. J. Keya, M. A. Hamid, M. M. Monowar, and M. S. Rahman, “A comprehensive review on fake news detection with deep learning,” *IEEE Access*, vol. 9, pp. 156 151–156 170, 2021.

[42] N. Capuano, G. Fenza, V. Loia, and F. D. Nota, “Content based fake news detection with machine and deep learning: a systematic review,” *Neurocomputing*, 2023.

[43] A. M. Mustafa and T. A. Rashid, “Kurdish stemmer pre-processing steps for improving information retrieval,” *Journal of Information Science*, vol. 44, no. 1, pp. 15–27, 2018.

[44] H. Q. Abonizio, J. I. De Morais, G. M. Tavares, and S. Barbon Junior, “Language-independent fake news detection: English, portuguese, and spanish mutual features,” *Future Internet*, vol. 12, no. 5, p. 87, 2020.

[45] B. Horne and S. Adali, “This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news,” in *Proceedings of the international AAAI conference on web and social media*, vol. 11, no. 1, 2017, pp. 759–766.

[46] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, “Lexicon-based methods for sentiment analysis,” *Computational linguistics*, vol. 37, no. 2, pp. 267–307, 2011.

[47] R. A. Monteiro, R. L. Santos, T. A. Pardo, T. A. De Almeida, E. E. Ruiz, and O. A. Vale, “Contributions to the study of fake news in portuguese: New corpus and automatic detection results,” in *Computational Processing of the Portuguese Language: 13th International Conference,*

PROPOR 2018, Canela, Brazil, September 24–26, 2018, Proceedings 13. Springer, 2018, pp. 324–334.

[48] J.-P. Posadas-Durán, H. Gómez-Adorno, G. Sidorov, and J. J. M. Escobar, "Detection of fake news in a new corpus for the spanish language," *Journal of Intelligent & Fuzzy Systems*, vol. 36, no. 5, pp. 4869–4876, 2019.

[49] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[50] C. Boididou, K. Andreadou, S. Papadopoulos, D. T. Dang Nguyen, G. Boato, M. Riegler, Y. Kompatsiaris *et al.*, "Verifying multimedia use at mediaeval 2015," in *MediaEval 2015*. CEUR-WS, 2015, vol. 1436.

[51] Z. Jin, J. Cao, H. Guo, Y. Zhang, and J. Luo, "Multimodal fusionwith recurrent neural networks for rumor detection on microblogs," in *Proceedings of the 25th ACM international conference on Multimedia*, 2017, pp. 795–816.

[52] A. Zubiaga, M. Liakata, and R. Procter, "Learning reporting dynamics during breaking news for rumour detection in social media," *arXiv preprint arXiv:1610.07363*, 2016.

[53] V. Slovikovskaya, "Transfer learning from transformers to fakenews challenge stance detection (fnc-1) task," *arXiv preprint arXiv:1910.14353*, 2019.

[54] T. Alhindi, S. Petridis, and S. Muresan, "Where is your evidence: Improving fact-checking by justification modeling," in *Proceedings of the first workshop on fact extraction and verification (FEVER)*, 2018, pp. 85–90.

[55] A. Pal, M. Pradhan *et al.*, "Survey of fake news detection using machine intelligence approach," *c*, vol. 144, p. 102118, 2023.

[56] A. G. Rohit Kumar Kaliyar and P. Narang, "Fakebert: Fake news detection in social media with a bert-based deep learning approach," *Springer Science+Business Media, LLC, part of Springer Nature 2021*, vol. 80, p. 100051, 2021.

[57] T. Chauhan and H. Palivela, "Optimization and improvement of fake news detection using deep learning approaches for societal benefit," *Information Management Data Insights*, p. 102118, 2021.

[58] S. E. Sorour and H. E. Abdelkader, "Afnd: Arabic fake news detection with an ensemble deep cnn-lstm model," *J. Theor. Appl. Inf. Technol.*, vol. 100, no. 14, pp. 5072–5086, 2022.

[59] 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.

[60] H. Gonen, S. Ravfogel, Y. Elazar, and Y. Goldberg, "It's not greek to mbert: inducing word-level translations from multilingual bert," *arXiv preprint arXiv:2010.08275*, 2020.

[61] S. Alqurashi, A. Alhindi, and E. Alanazi, "Large arabic twitter dataset on covid-19," *arXiv preprint arXiv:2004.04315*, 2020.

[62] Q. Li, H. Peng, J. Li, C. Xia, R. Yang, L. Sun, P. S. Yu, and L. He, "A survey on text classification: From traditional to deep learning," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 13, no. 2, pp. 1–41, 2022.

[63] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V. Le, "Xlnet: Generalized autoregressive pretraining for language understanding," *Advances in neural information processing systems*,vol. 32, 2019.

[64] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[65] X. Liu, P. He, W. Chen, and J. Gao, "Multi-task deep neural networks for natural language understanding," *arXiv preprint arXiv:1901.11504*, 2019.

[66] S. Kumar, S. Kumar, P. Yadav, and M. Bagri, "A survey on analysis of fake news detection techniques," in *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*. IEEE, 2021, pp. 894–899.

[67] R. Varma, Y. Verma, P. Vijayvargiya, and P. P. Churi, "A systematic survey on deep learning and machine learning approaches of fake news detection in the pre-and post-covid-19 pandemic," *International Journal of Intelligent Computing and Cybernetics*, vol. 14, no. 4, pp. 617–646, 2021.