Informatics Bulletin, Faculty of Computers and Artificial Intelligence, Helwan University
Published Online *7 **Issue 2, July 2025***
(https://fcihib.journals.ekb.eg)

A Comparative Analysis of Fake News Detection Techniques: A Case Study on Disinformation During the COVID-19 Pandemic

Samar M.Abd El-Mageed[1,2], Amal Elsayed Aboutabl[1], Ensaf Hussein Mohamed[1,3]
[1]Computer Science Department, Faculty of Computers and Artificial Intelligence, Helwan University, Cairo, Egypt
[2]Computer Science Department, Faculty of Computers and Information, Luxor University, Luxor, Egypt
[3]School of Information Technology and Computer Science, Nile University, Giza, Egypt
samar.mahmoud@fci.luxor.edu.eg, amal.aboutabl@fci.helwan.edu.eg, enmohamed@nu.edu.eg

***Abstract*—In the current era, social media serves as a highly influential platform for news dissemination, where information is readily accessible to the public without verification of its authenticity. Consequently, it has become an effective medium for the propagation of false information, enabling individuals to share unverified content that reaches a vast audience. This capability facilitates the creation and spread of misinformation, potentially misleading society for personal or corporate benefit. Fake news is often utilized to influence government policies and tarnish the reputations of individuals. In the medical field, the COVID-19 pandemic has resulted in a substantial rise in the use of social media platforms as sources of information. Unfortunately, this surge in usage has also contributed to the widespread dissemination of misinformation concerning healthcare and the pandemic. So, to facilitate the consequences of fake news on society, several research projects have been proposed to identify fake news with the best accuracy to protect its outcome. For this, we present a comparative literature of the fake text classification architectures in this paper and an overall analysis of the state of the art techniques and challenges that have been conducted on fake information and propaganda in Arabic and English texts, as well as the different approaches that have been taken to address this problem.**

***Index Terms*—Fake news, Deep learning, Arabic text, Transformers, COVID-19.**

I. INTRODUCTION

Fake information is the text that is verified to be untrue. [1]. This definition shows that fake news is false and is presented as if it were true. The storm of fake information extends the necessity of effective analytical tools for the ability to provide insight into the validity and reliability of online content [2]. Users of social media are especially vulnerable to the influence of fake news. They can be fooled by fake news on sensitive issues like healthcare, driven into echo chambers that deepen social divisions, and lose faith in real news sources. All this disinformation makes them easily manipulated and hinders informed decision-making. In May 2020, there were about 14,000 articles on PubMed on COVID-19, rendering it impossible to stay updated with recent publications.[3]. Social media serves as a powerful medium for knowledge dissemination in a short time. The reduced translation time for the knowledge provided by social media allows one to analyze the literature in real time. An example of this is the recent online debate surrounding studies on hydroxychloroquine and remdesivir. Social media is an ideal and short way for everyone to spread and manipulate false news. According to Facebook's reports, manipulative actions by malicious actors constituted less than 0.1% of the public content shared on the platform [4], [5]. In 2008, fake news about Steve Jobs' health caused a huge change in the stock market performance of Apple Inc [6]. For example, research indicates that approximately 19 million bot accounts posted tweets in favor of both Clinton and Trump during the 2016 US presidential election [7], which ideally shows that social media helps hugely in the creation and spread of fake news. The fake news classification task refers to the identification process of classifying text as false or misleading information that is presented. Fake detection acts as a digital antidote, identifying and neutralizing misleading information that circulates on social media. Through automated investigations and human fact-checkers to expose the deceptive use of fake news, dismantling its potential for manipulation and disinformation. This important process helps to restore trust in legitimate resources and provides a conducive online environment for informed discussion.

This approach used various techniques, such as linguistic analysis and machine learning models, to identify characteristic patterns of fake news. The essential objective of fake text detection is to help individuals and organizations recognize its potentially harmful effects, which can extend beyond personal impacts to broader societal consequences. It is hard to detect fake information. However, recent advances in artificial intelligence, especially in computational capabilities and big data analysis, have demonstrated significant potential to tackle the challenges of fake news detection [8], [9]. Artificial intelligence plays a vital role in different areas of human life. Machine learning (ML) and deep learning (DL) techniques have been extensively applied to the detection of fake news. Researchers have used methods such as logistic regression, Naive Bayes (NB), Support Vector Machines (SVM), and Decision Trees (DT) to effectively address this problem [10].

Convolutional Neural Networks (CNN) and DNN have also been employed to detect fake news, achieving remarkably high accuracy in the process [11], [12]. Driven by these observations, we provide an extensive review of advanced techniques for fake news classification.

Many people use social networks due to their accessibility and user-friendly nature, which has facilitated the rapid spread of fake news, regardless of its credibility. Social networks often exploit critical situations, such as the COVID-19 pandemic and the US presidential elections, to negatively influence communities. Fake content can have severe consequences across various sectors, including finance, politics, and sports. Research indicates that misinformation and false health-related content on social media during pandemics, health emergencies, and humanitarian crises can lead to significant mental, social, political, and economic harm. The present surveys analyze the different AI techniques. We implemented a few techniques and discussed their results. This research, conducted from 2019 to 2023, focuses on the identification of fake news within the Arabic medical domain.

The main contributions of this research are as follows:

- We provide an extensive review of the classification methods in AI for fake news, along with an exploration of various datasets used for fake news detection.
- We discuss the research challenges associated with the latest AI techniques developed for fake news classification.

We implement the Bi-LSTM, LR, and SVC algorithms for fake news classification. The Naive Bayes (NB) algorithm is effective for high-dimensional datasets, offering fast performance with minimal tunable parameters. LSTM is chosen due to its reputation as a state-of-the-art approaches, while SVC excels in handling large datasets. The Performance Assessment Section provides a detailed analysis of the results from these methods. Figure 1 shows the construction of this review. Section 2 establishes a foundation by providing an overview of fake news. Section 3 shows that the datasets employed for analysis, outline their characteristics and suitability for fake news classification. Section 4 explores a range of fake news classification techniques. Section 5 specifically addresses the challenges associated with disinformation detection in the healthcare sector, highlighting unique considerations. Section 6 presents a comprehensive performance evaluation of the implemented techniques, providing valuable insights into their effectiveness. Finally, Section 7 summarizes the main findings, along with potential future research directions in this dynamic field.

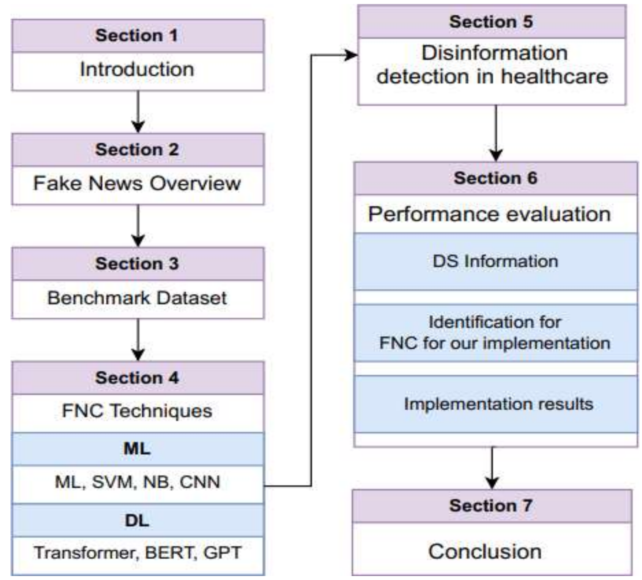


Fig. 1. Survey Structure

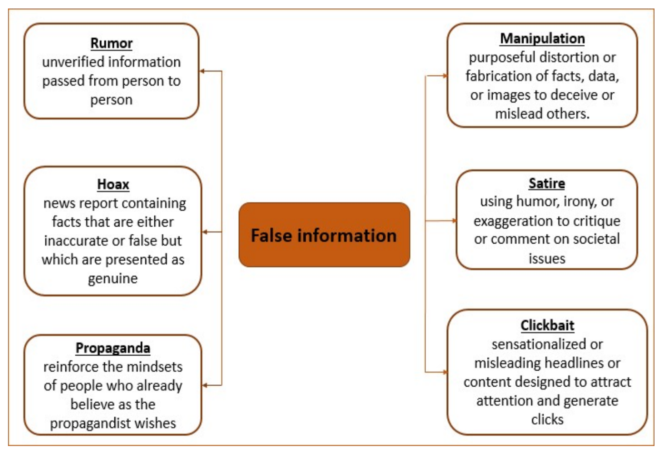


Fig. 2. False information types

2. Fake News Overview

Fake news is fabricated information masquerading as real news, often distributed online to mislead or influence people's minds for profit. There are several types of disinformation represented in Figure 2

- Rumor: it is defined as a sentence and its truth value is valid [13]. The challenges of rumor detection on social media have garnered significant attention. Most publications have focused on detecting rumors through the superficial features of messages, including the characteristics and content of the posters. There are a lot of different publications about rumors that have different definitions [14, 15, 16].
- Hoax: It is a deliberate attempt to deceive someone by presenting something as real. It can also be a made-up story, a joke, or a clever sketch. The confusion is often spread as a joke, an embarrassment, or something to say about a particular topic. When the hoax increases on social media, it heightens the risk of mass panic, as society becomes confused between lies and truth, or in other words, between deception and reality. Many research papers describe the detection of deception on social media [17, 18, 19].

- Propaganda: It is the sharing of information (such as posts, videos, or images) to influence your opinion on a topic, usually through the use of emotionally charged, misleading information about people, or the actual system of hiding it. Basically, trying to sway you in a certain direction without giving you all the facts. [20, 21, 22] describes different techniques, such as supervised and unsupervised algorithms, to detect propaganda on social media.
- Manipulation: Fake news manipulates text by twisting facts, using biased language, or inventing sources to deceive readers. [23] describes manipulation detection in the era of COVID-19.

3. Benchmark Datasets

In this part, we will talk about various Arabic and English datasets used in different studies. Standard datasets have been used for training and testing. One of the challenges in detecting fake news is the lack of a labeled dataset with true and false labels, as well as the scarcity of large-scale datasets. Datasets can vary based on their language, mode of use, and size. Therefore, based on these characteristics, we categorize these datasets in Table 1. The LIAR dataset contains around 12,800 labeled short statements from the Politifact website, which were manually fact-checked. Each statement is labeled with one of six labels. The dataset includes metadata like the subject, speaker, and context, which provides a substantial source of news for models to learn from. A Hierarchical Attention Network (HAN) is proposed for fake news detection, using the LIAR dataset in [24]. The network uses hierarchical attention mechanisms to focus on the key words and sentences in a news article. In [25], the use of the BERT model for fake news detection on the LIAR dataset is explored. It examines how contextual embeddings can improve classification accuracy. An extension of the Fake News Net dataset, it includes not just the news content but also the social context, including user interactions and the structure of social networks. This additional data enables researchers to explore how fake news spreads on social media and how social context can improve detection accuracy. In [26], it is shown that integrating news content with social context, including user interactions and social network structures, significantly enhances the accuracy of fake news detection using this dataset. In [27], it highlights how GNNs can be used to model social interactions and enhance the interpretability of fake text classification systems using the same dataset. [28] demonstrates that combining multiple modalities, including social context, can significantly enhance fake news detection performance using Fake News Net. The COVID-19 Fake News Dataset (CoAID) is prepared to detect fake information related to the COVID-19 pandemic. It includes news articles, social media posts, and user engagements. The dataset is particularly useful for studying misinformation during global health crises, with labeled articles on COVID-19 from multiple sources. In [29], a deep learning approach for detecting fake news related to COVID-19 using the CoAID dataset is presented. The authors compare various models, including LSTM, CNN, and transformers, to evaluate their effectiveness in this specific domain. In [30], it proposes a multi-modal approach that integrates textual, visual, and social context information to detect COVID-19 misinformation. The CoAID dataset is used to train and validate the proposed models. AraFake is a dataset designed specifically for Arabic fake news detection. It contains thousands of articles composed from various sources, labeled as either true or not.

The dataset includes news articles from diverse domains, such as politics, health, and entertainment. The labeling process often involves manual verification and cross-referencing with trusted news sources. AraFake is utilized to train machine learning models for binary classification tasks, distinguishing between fake and real news. It also supports research into the linguistic features specific to Arabic that might influence fake news detection. In [31], it presents AraNet, a deep learning model specifically designed for Arabic fake news detection. The model leverages the AraFake dataset for training and evaluation, demonstrating significant improvements over traditional methods. In [32], it introduces the AraFake dataset and provides benchmark results using several machine learning approaches, including SVM, logistic regression, and deep learning techniques.

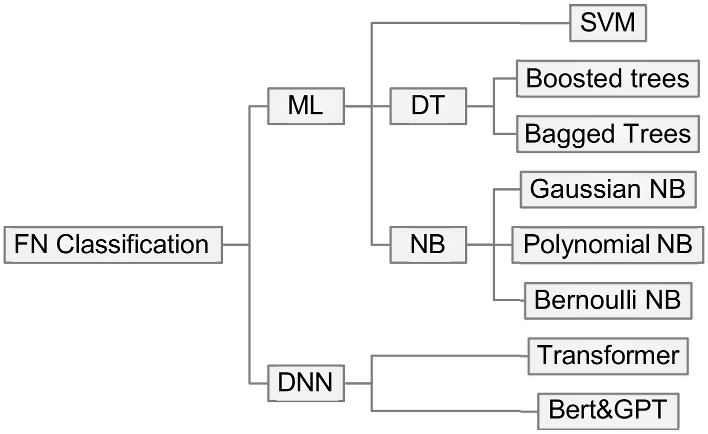
The ArCOV19-Rumors dataset focuses on misinformation related to COVID-19 in Arabic. It includes social media posts, news articles, and rumors that circulated during the pandemic. The dataset contains a broad area of topics, from health-related misinformation to conspiracy theories. Each entry is labeled based on its veracity, with categories such as ”true”, ”false”, and ”uncertain”. Researchers use ArCOV19-Rumors to analyze how COVID-19 false information spreads in Arabic-speaking communities and to develop models for detecting such content. In [33], it proposes a multi-feature approach for detecting COVID-19 misinformation on Arabic social media platforms. The approach integrates linguistic, content-based, and social network features and is evaluated using the ArCOV19-Rumors dataset. In [34], it explores the application of transfer learning techniques and data augmentation to enhance the performance of COVID-19 misinformation detection models in Arabic. The ArCOV19-Rumors dataset is used for training and testing the proposed models.

TABLE 1: Details of Arabic and English dataset

Language	**Dataset**	Modality	**Size**	Type
English	Fake News	Text	20,8K	News articles
	FacebookHoax	Text	15,5K	Scientific news
	LIAR	Text	12.8K	Political statements
	ISOT	Text	12.6K	News articles
	Twitter	Text & Images	992	Fact checking and claim
	KaggleFN	Text	13K	Political news
	FakenewsNet	Text	5K	News articles, Social media data
Arabic	AFND	Text	606912	News articles
	Arabic news stance (ANS)	Text	3K	News articles
	Arabic COVID-19 Rumor	Text	30K	News articles
	AraNews	Text	108K	News articles

4. Techniques of Fake News Classification

This section presents a taxonomy for fake news identification as shown in the following chart.



FNC techniques

We analyze various techniques that utilize machine learning and deep learning models for fake news classification. Figure 3 describes identification steps in the process of fake news classification.

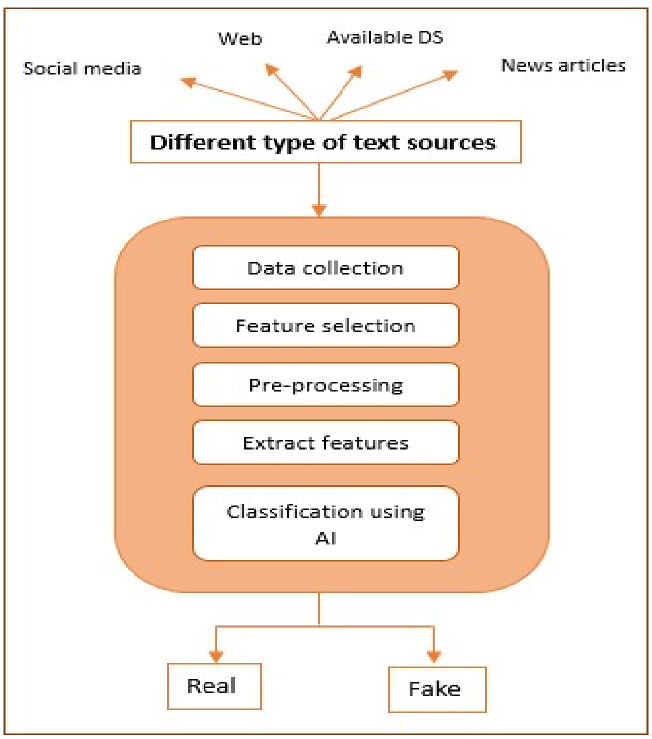


Fig. 3. Identification steps of FND

A. Machine Learning

ML plays an essential role in fake news classification. By training on data, these algorithms can learn to recognize patterns in text, user behavior, and sharing habits and these systems can then analyze new features and verify them more accurately. In this section, we explore the supervised learning models, including SVM, KNN, DT, NB, and LR, which have been employed to classify fake news.

TABLE 2: Different SVM Approaches for FNC

Author	Year	Dataset used	Pre-processing steps	Results
G. Hussain and et al [35]	2020	Bangla news from social media	Removing -The numerical values, -Punctuation marks, -Special symbols. TF-IDF	93.32%
K.Yazdi and et al [36]	2020	BuzzFeedNews BS Detector LIAR	K mean F5	95.3% 90.7% 91.7%
A.Mahabub and et al [37]	2020	Data used by Wang	Removed special characters	80%
A.Elmogy and et al [38]	2021	Yelp dataset	Tokenization Stop word cleaning Lemmatization TF-IDF	87.8%

TABLE 3: Different DT Approaches for FNC

Author	Year	Dataset used	Pre-processing steps	Results
R.Jehad and et al [40]	2020	fake news articles	Removing -unnecessary special characters -numbers -letters -spaces -stop words TF-IDF	89.11%
T.Dinesh and et al [41]	2021	Fake political news	Removing special characters	99.6%
N.Krishna and et al [42]	2022	fake news articles	Removing special characters	97.67%

1) Support Vector Machine: Table 2 represents the results of several studies that have used SVM for fake news classification. The studies used different datasets and preprocessing steps, but all of them achieved relatively high accuracy. The table also illustrates that the preprocessing steps can significantly impact the accuracy of SVM for fake news detection. For example, the study by K.Yazdi et al. achieved an accuracy of 95.3% on the BuzzFeed News dataset using K-means, F5, and LINEAR features. Overall, the table suggests that SVM is a promising machine-learning algorithm for fake news detection. Table 3 summarizes the findings of several studies on fake detection in Arabic text using SVM.

2) Decision Tree (DT): There are two types of DT:

1) Boosted Trees: Builds an ensemble of trees by training each new model to focus on the instances that were previously misclassified by the earlier models.
2) Bootstrap Aggregated (Bagging): Constructs decision trees by repeatedly re-sampling the training data with replacement, and then combines the predictions of these trees through voting during the prediction phase.

Table 3 shows the results of fake news article detection using decision tree algorithms applied in three different studies. Each study used a different dataset of fake news articles. All studies preprocessed the text data by removing unnecessary special characters, numbers, letters, spaces, and stop words.
In [39] employs a decision tree algorithm to detect fake news articles in Arabic. The experiment utilized an Arabic Corpus of Fake News. Preprocessing techniques included removing extraneous elements like URLs, foreign language words, and duplicate comments. The decision tree achieved an accuracy of 95.5% in identifying fake news articles. While this finding suggests promise for decision trees in Arabic fake news detection, we find that there are a few studies, and further research is needed to determine generalizability across different datasets and pre-processing methods.

TABLE 4: Different NB Approaches for FNC

Author	Year	Dataset used	Pre-processing steps	Results
F.Adiba and et al [44]	2020	Fake and Real news dataset	Tokenization Stemming TfIdfVectorization CountVectorization	92%
N.Yuslee and et al [45]	2021	News articles	Removing -the stop words lemmatization TF-IDF	94%
A.Yerlekar and et al [46]	2021	Medical news articles	Remove stopwords Countvectorization TF-IDF	80%
S.Senhadji and et al [47]	2022	Fake news from kaggle	Remove -special characters -spaces -URLs Lemmatization	92%

TABLE 5: DL Approaches for FNC

Author	Year	Dataset used	Pre-processing steps	Results
P.Hajek and et al [54]	2020	datasets from Cornell University	n-grams, word embeddings lexicon based	88%
N.Aslam and et al [55]	2021	LIAR	Explicit feature extraction	89%
R.Kaliyar and et al[56]	2021	BuzzFeed and PolitiFact	Mapping data to word embedding	91%
M.Berrahal and et al [57]	2023	CelebA-HQ	Removing special characters and spaces	99.4%

TABLE 6: CNN Approaches for FNC

Author	Year	Dataset used	Pre-processing steps	Results
Q.Hu and et al [58]	2019	LIAR NewsFN	N-gram	91.6% 92%
H.SALEH and et al [59]	2021	Kaggle dataset + FakeNewsNet + FA-KESS	Remove special characters Tokenization stemming	98.6%

3) Naive Bayes (NB): There are three types of NB distributions:

1) Gaussian NB: Used when the features have a continuous value range.
2) Polynomial NB: Primarily applied in document classification problems, where the words' frequency in the document is applied as the distribution.
3) Bernoulli NB: As to Polynomial Naive Bayes, the features representing the inputs are boolean variables. This algorithm is commonly used in applications such as recommendation systems, spam filtering, and sentiment analysis.[43].

Table 4 represents the fake news article detection results using NB algorithms applied in four different studies. Each study used a different dataset of fake news articles.

B. Deep Neural Network

Deep neural networks (DNNs) are emerging as powerful tools in the fight against fake news. Their complex architecture allows them to analyze various aspects of text data that can be crucial for identifying deception. For instance, CNNs within a DNN can effectively capture local patterns in language, like the use of inflammatory words or unusual grammatical structures. LSTMs, another common component of DNNs, excel at understanding the sequential nature of language. This enables them to analyze the flow of information and identify inconsistencies or illogical connections within the text, which are often hallmarks of fake news. By combining these capabilities, Deep neural networks can achieve an accurate understanding of the content and context of news articles, leading to more precise detection of fake news.

The combination of CNNs and LSTM networks has stood out as a popular and effective way, demonstrating high accuracy in numerous research efforts [48, 49, 50]. This success can be attributed to CNNs' ability to capture local features within text data, while LSTMs excel at modeling the sequential nature of language, allowing for a holistic grasp of context and meaning. within an article. The combination of (CNNs) and (LSTM) networks is gaining significant traction in the domain of fake news detection for the Arabic also. These hybrid networks have demonstrated high accuracy rates in numerous studies. This allows them to delve deeper into the context and true meaning of Arabic news articles [51, 52, 53]. Table 5 represents some studies.

1) Convolutional Neural Network (CNN): This algorithm is considered the most widely used among other supervised learning techniques. To train this model effectively, a large volume of data is required to fully leverage its capabilities. Table 6 represents some studies that applied CNN in fake news detection.

2) Transformers: In 2017, Vaswani et al. [60] introduced the Transformer, a groundbreaking encoder-decoder architecture. Unlike previous models that processed text sequentially, the Transformer can handle all input tokens (words) at once. It treats the input sequence as a collection of tokens, without relying on their order. To understand relationships between words, the Transformer uses a mechanism called "self-attention." However, since order information is absent, another step is crucial: positional encoding.

This special technique, applied before the first processing layer, ensures that the same word appearing in different positions within a sentence has a distinct representation. Positional encoding injects information about the relative location of words, compensating for the lack of inherent order in the Transformer's input. Figure 4 shows The multi-head attention layer used in the Transformer architecture.

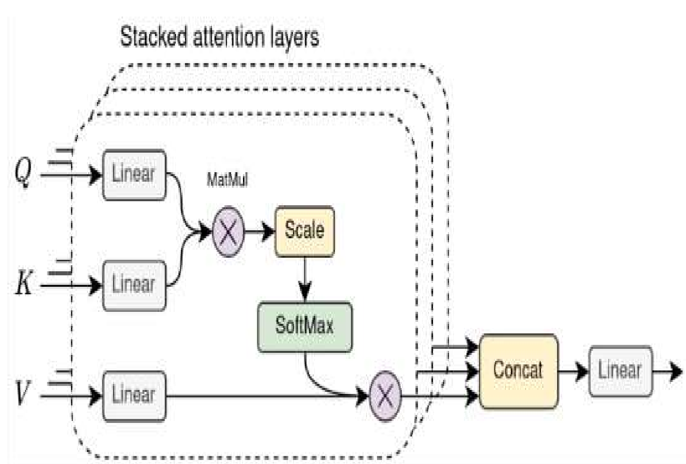


Fig. 4. The multi-head attention layer used in the Transformer architecture [61]

3) BERT and GPT:

- BERT
  Building upon the success of GPT, Bidirectional Encoder Representations from Transformers (BERT) [62] introduced a key difference: bidirectionality. Unlike GPT, which predicts words sequentially, BERT leverages a multi-layered Transformer encoder to analyze text in both directions. BERT's training method also sets it apart. Instead of next-word prediction, BERT employs a "masked language model" (MLM) technique. Here, a portion of the input text is masked, and the model predicts the masked words based on the surrounding context. This strategy is essential for BERT to achieve bidirectionality, allowing it to understand the relationships between masked words and both their preceding and following words. An additional training task for BERT involves next sentence prediction (NSP). The model is presented with two sentences and must predict whether the second one follows the first logically. This helps BERT learn relationships between sentences. Interestingly, adapting BERT for various tasks is straightforward. Researchers have achieved impressive results in classification by fine-tuning a pre-trained BERT model. This involves passing the encoded text representations from BERT through a simple neural network for task-specific adjustments.
- GPT
  While the original Transformer architecture excels at language modeling, some researchers believe it might learn redundant information during encoding and decoding. They propose that using only the encoder or decoder part could achieve similar performance with smaller models [63]. Building on this idea, the Generative Pre-trained Transformer (GPT) [64] utilizes a decoder-only architecture with multiple stacked Transformer decoder layers. The core concept behind GPT is similar to language modeling-predicting the next word in a sequence. Since it's an autoregressive model, GPT's predictions rely on the preceding context (left-to-right in the original architecture). This means it's not bidirectional, unlike models like BiLSTMs. The decoder-only architecture offers advantages for handling longer text sequences, making GPT well-suited for generative tasks like abstractive text summarization and question answering. The GPT authors also provide adaptations for various downstream tasks, including classification.

5. Disinformation Detection Challenges in Healthcare

Health literacy and understanding of medical terms are important aspects in the area of computer science, particularly in areas where technology intersects with healthcare and medical informatics. Natural Language Processing (NLP) is a branch of computer science dedicated to enabling interactions between computers and human languages. In healthcare, NLP approaches are applied to analyze and understand medical sentences such as research papers, electronic health records, patient forums, and clinical notes. Understanding medical terminology is crucial for accurately extracting information and extracting meaningful insights from these texts. False information shared on social media platforms poses a significant problem. And if this news is related to an important and vital event. The first and most significant misinformation in the field of public health is the false belief that the MMR vaccine causes autism. Andrew Wakefield, along with 12 colleagues, published a case series in *The Lancet* that proposed a potential link between the MMR vaccine and behavioral regression in children [65].

In 2020, several countries, including the United Kingdom, Greece, Venezuela, and Brazil, lost their status of measles elimination. In the field of cardiology, fake news is also prevalent. Social media spread a significant amount of misinformation regarding the potential cancer-causing effects of antihypertensive drugs, leading many patients to discontinue the use of certain proven beneficial medications [66]. The development of vaccines has been a comfort among all people. So, many governments have made efforts to administer vaccines to their populations. However, the distribution of these vaccines faces challenges due to people's fears. COVID-19 vaccines have sparked widespread discussions on social media, influencing public opinions about vaccination. Additionally, privacy and security concerns regarding personal health information on social media are substantial and require careful attention. Here are some common issues and risks associated with sharing health information on social media platforms: Sharing personal health information on social media can increase the risk of unauthorized access by malicious actors. This can lead to identity theft, fraud, or other misuse of sensitive health data. Social media platforms may experience data breaches, resulting in the exposure of personal health information to unauthorized individuals. Breaches can occur due to security vulnerabilities or attacks targeting the platform's infrastructure, and Lack of data accuracy and context: Health information shared on social media may lack accuracy, context, or appropriate medical expertise. Relying on such information for medical decision-making can lead to misinformation, misinterpretation, or inappropriate treatment choices. Table 7 shows different datasets used in fake news detection in healthcare. Table 8 shows studies in fake detection in COVID-19.

TABLE 7: Datasets used in fake detection in COVID-19 news

Language	Dataset	Size
English	Covid-19 fake news	10,700 posts and article
	Covid-19 vaccines	15,465,687 tweet
	COVID-19 dataset	6420 sample
Arabic	AraCovid19-MFH	10,282 tweet
	Arabic COVID-19 Rumor	9.4k tweet
	ArCovidVac	10k tweet

6. Performance Evaluation and Comparisons

To evaluate modern methods used for detecting fake news, we used data containing real tweets and fake ones. This dataset

TABLE 8: Relative Comparisons Approaches for Fake News Classification in COVID-19

Author	Year	Technique	Accuracy
English			
E.Shushkevich and et al [67]	2021	BERT	94%
S.M.Shifath and et al [68]	2021	NN	94%
M.Samadi and et al [69]	2021	ROBERT + MLP	97%
B.Al-Ahmad and et al [70]	2022	MLP	-
Arabic			
F.Haouari and et al [71]	2020	AraBERT+MARBERT	75%
M.Seghir and et al [72]	2021	AraBERT+mBERT	98%
H.Mubarak and et al [73]	2022	AraBERT	83%

is the first large-scale collection of Arabic tweets manually annotated for the COVID-19 vaccination campaign. It spans several countries in the Arab region and includes multiple layers of annotations to enhance content understanding.

A. Dataset Information

We have used the largest manually annotated Arabic tweet dataset, ArCovidVac [74], for the COVID-19 vaccination campaign, covering many countries within the Arab world. This data is enriched with special layers of annotation and informativeness: classifying tweets by their relevance (more vs. less informative). Content types: identifying the specific type of information in the tweet (e.g., advice, rumors, restrictions, or confirmed news). Sentiment: categorizing tweets based on their attitude toward vaccination (pro-vaccine, neutral, or anti-vaccine). We did an in-depth analysis of the stance closer to the vaccination layer to classify tweets as either fake or not. In the data, each statement is annotated with its text, category, and stance. The number of samples isn't balanced, so we apply some EDA techniques to balance the dataset. Table 9 represents statistics of the dataset after the EDA process.

TABLE 9: Used dataset statistics before and after augmentation

	Real	Fake	Total
Without EDA	7962	1391	9353
With EDA	7962	7815	15777

The dataset is split into training and testing sets, with 70% allocated for training and the remaining 30% for testing. Feature extraction techniques, including TF-IDF and count vectorizer, are applied to the dataset using Python NLP packages. During the pre-processing phase, stop words are removed to improve accuracy. Figure 5 shows data after and before the cleaning steps.

B. Identification of fake news classification for our implementation

Here's a breakdown of the key steps involved:

- Pre-processing steps: This stage tackles various tasks to preprocess and prepare text data for additional processing. It includes removing special characters like punctuation marks, extra spaces, and URLs. Additionally, normalization of diacritics is performed. Here are some key aspects of data cleaning for a used dataset:

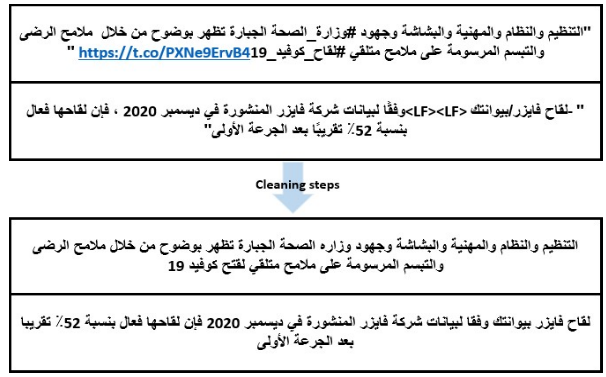


Fig. 5. Data after and before the cleaning steps

  - Diacritics: Remove or normalize diacritics (harakat) based on your specific use case.
  - Tashkeel: Similar to diacritics, consider handling tashkeel (vowel points) depending on your needs.
  - Whitespace: Normalize whitespace characters for consistency.
  - Text normalization: Convert ligatures, special characters, and named entities to standard forms.
  - Stop word removal: Remove common words with little semantic meaning depending on your task.

- Text Encoding: The text is transformed into a numerical representation that machines can understand. This is typically done using two methods:
- TF-IDF: This technique assigns weights to words based on the frequency in the document and its rarity across the entire text corpus. Words that appear frequently in a single document but rarely overall receive higher weights, indicating their significance for that specific document.
- SVC, Bi-LSTM, LR: these represent different machine learning algorithms used for feature extraction. They analyze the text data to capture essential characteristics and patterns that can be employed for various tasks like classification or prediction. Figure 6 represents flow for identification of implementations.
- BertForSequence Classification. The overall Structure is

  - Input: The model takes a sequence of tokens as input, represented by their integer indices.
  - BERT Model: Pre-trained transformer-based language model that captures rich contextual information from the input text.
  - Key Components: We used the Dropout for Regularization technique that prevents overfitting. And a classifier layer that converts the learned representation into class probabilities.

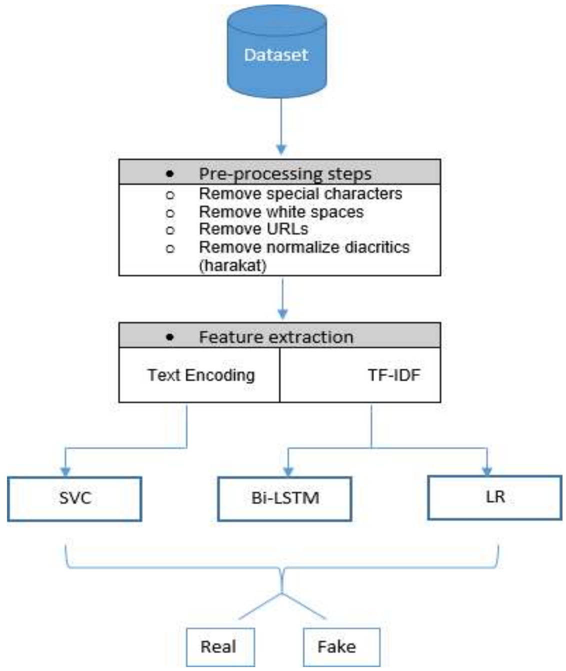


Fig. 6. Flow for identification of implementations

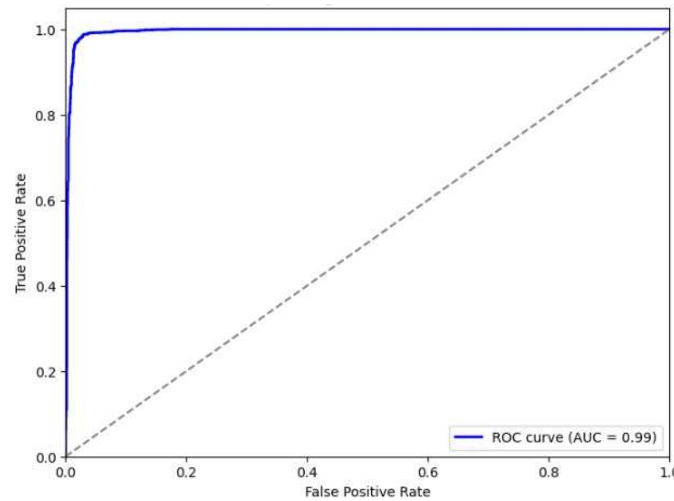
TABLE 10: Ensemble learning-based approaches for fake news classification

Approach	F1	Precision	Recall	Accuracy	ROC-AUC Score
SVC	0.9729	0.9736	0.9729	0.9729	0.9926
Bi-LSTM	0.9624	0.9634	0.9624	0.9624	0.9627
LR	0.9427	0.9454	0.9454	0.9427	0.9900
BERT	0.9472	0.9474	0.9475	0.9500	0.9499

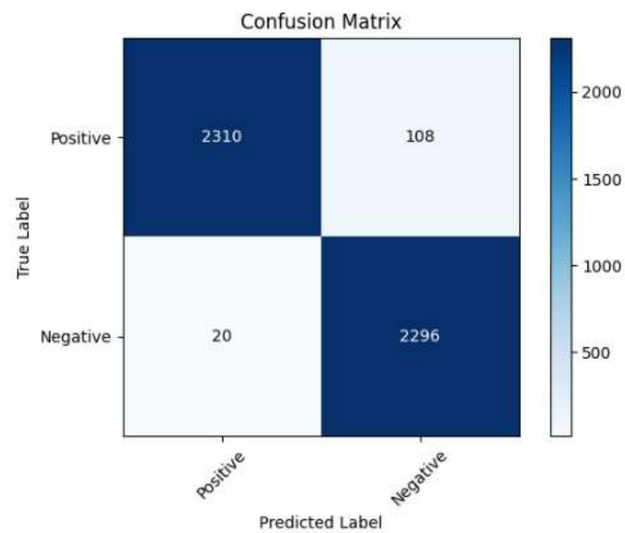
C. Implementation Results

Figures 7, 8, 9, and 10 represent ROC curves and confusion metrics for SVC, Bi-LSTM, LR, and BERT, respectively. Table 10 shows a relative comparison of these approaches. The experimental results demonstrate that balancing the dataset through EDA significantly improved classification performance across all models. The evaluation metrics show that SVC (Support Vector Classifier) outperforms all other models in this study, yielding the highest values for f1, precision, recall, accuracy, and ROC-AUC scores. The SVC's ability to provide a balanced and high-quality classification suggests that it is particularly effective at distinguishing between classes. Bi-LSTM and BERT represent strong deep learning alternatives that can capture more intricate patterns, especially for sequential and textual data. While their performance in traditional classification metrics is strong, there is room for improvement in generalization and training efficiency.

All experiments were conducted on a machine with an Intel Core i7 processor, 16GB RAM, and an NVIDIA GTX 1060 GPU using Python 3.8 and Jupyter Notebook. To ensure reproducibility, we fixed random seeds for NumPy, TensorFlow, and scikit-learn.



(a) ROC curve

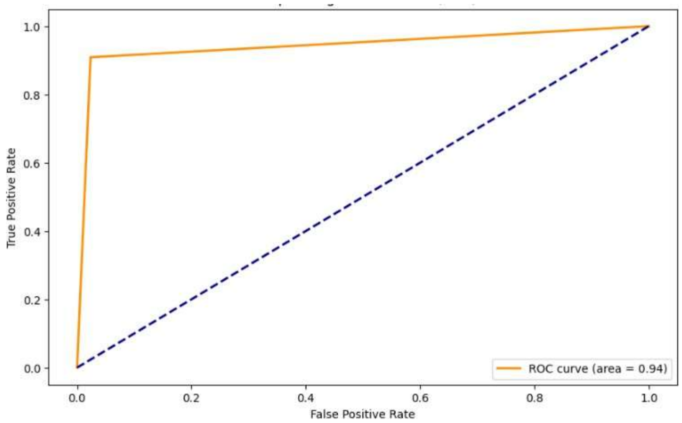


(b) Confusion matrix

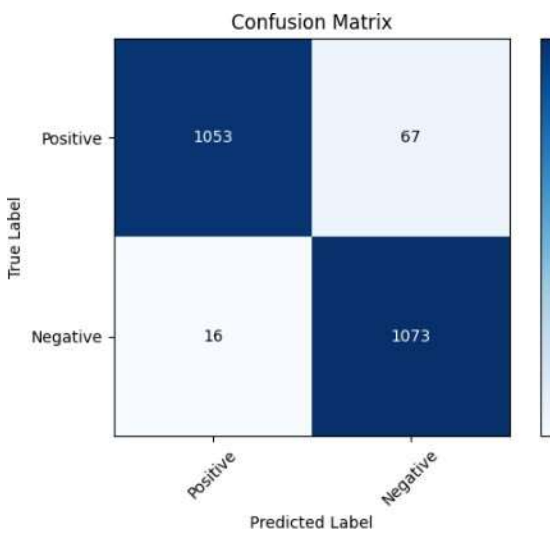
Fig. 7. ROC curve and Confusion matrix for SVC

7. Conclusion

Fake news detection (FND) techniques play a crucial role in combating the spread of misinformation. particularly in the healthcare domain. With the rapid growth of social media and the increasing impact of fake news on public health, developing effective FND techniques is of paramount importance. State-of-the-art FND techniques incorporate a combination of natural language processing, machine learning algorithms, network analysis, and integration with external sources. These techniques leverage textual features, semantic analysis, sentiment analysis, and machine learning models. to identify patterns of disinformation and distinguish between reliable and misleading information. To overcome these challenges, future directions and recommendations include the development of large-scale and diverse datasets, incorporation of multimodal information, improved explainability and interpretability, real-

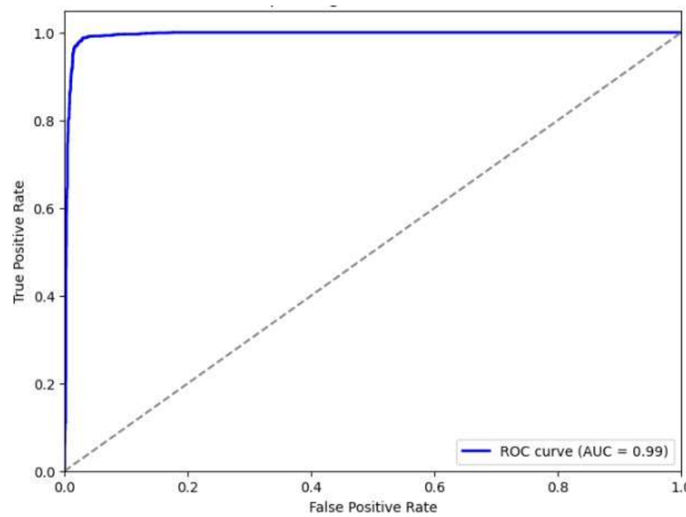


(a) ROC curve

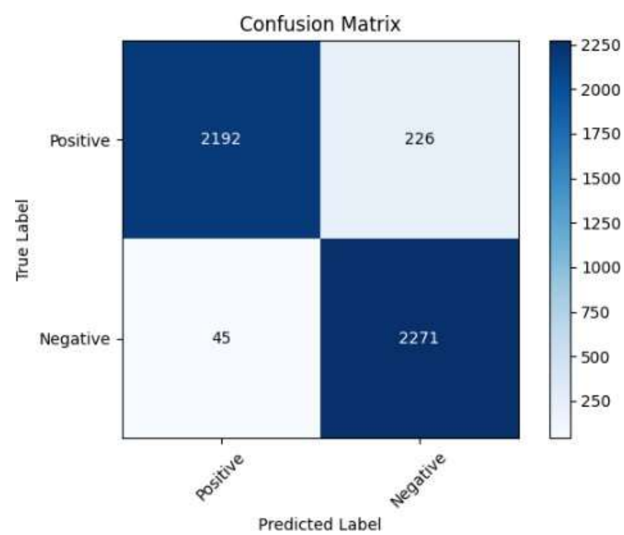


(b) Confusion matrix

Fig. 8. ROC curve and Confusion matrix for Bi-LSTM



(a) ROC curve



(b) Confusion matrix

Fig. 9. ROC curve and Confusion matrix for LR

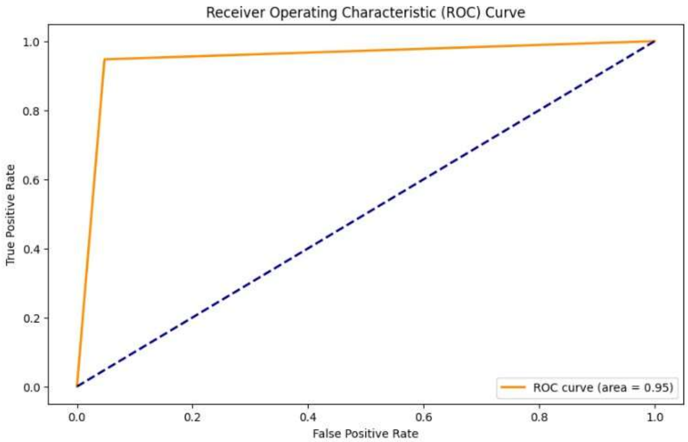
time and scalable detection systems, collaboration with healthcare professionals, user education and awareness, and policy and regulatory interventions. By addressing with these recommendations, researchers and stakeholders can enhance the effectiveness of FND techniques. Promote health literacy and protect public health from the harmful effects of fake news. Continued advancements in FND techniques will be vital in ensuring the availability of accurate and trustworthy healthcare information in the digital age. While significant progress has been made in fake news detection for English-language content, research in the Arabic medical domain remains relatively limited. The limited number of studies in this area highlights a significant gap in the literature, especially considering the growing spread of misinformation in healthcare-related content. Bridging this gap is essential to ensure the accurate and reliable dissemination of medical information within Arabic-speaking communities.

REFERENCES

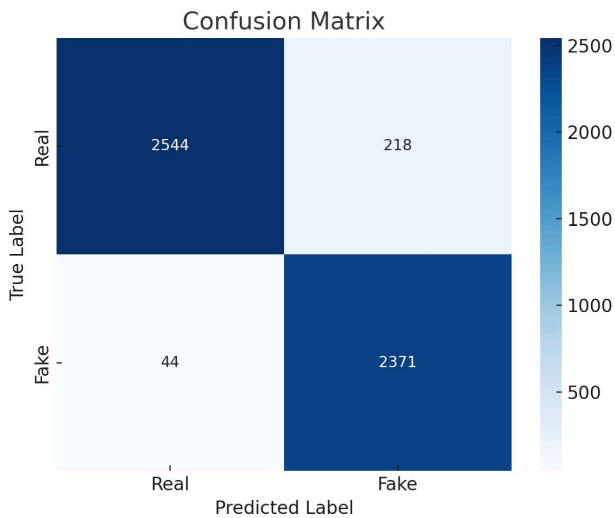
[1] P. Patwa, M. Bhardwaj, V. Guptha, G. Kumari, S. Sharma, S. Pykl, A. Das, A. Ekbal, M. S. Akhtar, and T. Chakraborty, "Overview of constraint 2021 shared tasks: Detecting english covid-19 fake news and hindi hostile posts," in *Combating Online Hostile Posts in Regional Languages during Emergency Situation: First International Workshop, CONSTRAINT 2021, Collocated with AAAI 2021, Virtual Event, February 8, 2021, Revised Selected Papers 1*. Springer, 2021, pp. 42–53.

[2] V. Pe´rez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, "Automatic detection of fake news," *arXiv preprint arXiv:1708.07104*, 2017.

[3] M. Gottlieb and S. Dyer, "Information and disinformation: social media in the covid-19 crisis," *Academic emergency medicine*, vol. 27, no. 7, p. 640, 2020.



(a) ROC curve



(b) Confusion matrix

Fig. 10. ROC curve and Confusion matrix for BERT

[4] J. Weedon, W. Nuland, and A. Stamos, "Information operations and facebook," *Retrieved from Facebook: https://fbnewsroomus. files. wordpress. com/2017/04/facebook-and-information-operations-v1. pdf*, 2017.

[5] J. Vora, S. Tanwar, S. Tyagi, N. Kumar, and J. J. Rodrigues, "Home-based exercise system for patients using iot enabled smart speaker," in *2017 IEEE 19th international conference on e-health networking, applications and services (Healthcom)*. IEEE, 2017, pp. 1–6.

[6] E. Schonfeld, "Citizen journalist hits apple stock with false (steve jobs) heart attack rumor," *TechCrunch*, 2008.

[7] K. Stahl, "Fake news detection in social media," *California State University Stanislaus*, vol. 6, pp. 4–15, 2018.

[8] J. Clark, "Why 2015 was a breakthrough year in artificial intelligence. bloomberg," *Com https://www. bloomberg. com/news/articles/2015-12-08/why-2015-was-a-breakthroughyear-inartificial-intelligence*, 2015.

[9] R. Gupta, M. M. Patel, A. Shukla, and S. Tanwar, "Deep learning-based malicious smart contract detection scheme for internet of things environment," *Computers & Electrical Engineering*, vol. 97, p. 107583, 2022.

[10] K. Sheth, K. Patel, H. Shah, S. Tanwar, R. Gupta, and N. Kumar, "A taxonomy of ai techniques for 6g communication networks," *Computer communications*, vol. 161, pp. 279–303, 2020.

[11] K. Shu, D. Mahudeswaran, and H. Liu, "Fakenewstracker: a tool for fake news collection, detection, and visualization," *Computational and Mathematical Organization Theory*, vol. 25, pp. 60–71, 2019.

[12] P. Thakkar, K. Varma, V. Ukani, S. Mankad, and S. Tanwar, "Combining user-based and item-based collaborative filtering using machine learning," in *Information and Communication Technology for Intelligent Systems: Proceedings of ICTIS 2018, Volume 2*. Springer, 2019, pp. 173–180.

[13] S. Hamidian and M. T. Diab, "Rumor detection and classification for twitter data," *arXiv preprint arXiv:1912.08926*, 2019.

[14] H. Guo, J. Cao, Y. Zhang, J. Guo, and J. Li, "Rumor detection with hierarchical social attention network," in *Proceedings of the 27th ACM international conference on information and knowledge management*, 2018, pp. 943–951.

[15] Q. Zhang, S. Zhang, J. Dong, J. Xiong, and X. Cheng, "Automatic detection of rumor on social network," in *Natural Language Processing and Chinese Computing: 4th CCF Conference, NLPCC 2015, Nanchang, China, October 9-13, 2015, Proceedings 4*. Springer, 2015, pp. 113–122.

[16] M. Al-Sarem, W. Boulila, M. Al-Harby, J. Qadir, and A. Alsaeedi, "Deep learning-based rumor detection on microblogging platforms: a systematic review," *IEEE access*, vol. 7, pp. 152 788–152 812, 2019.

[17] A. Zubiaga and A. Jiang, "Early detection of social media hoaxes at scale," *ACM Transactions on the Web (TWEB)*, vol. 14, no. 4, pp. 1–23, 2020.

[18] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *ACM SIGKDD explorations newsletter*, vol. 19, no. 1, pp. 22–36, 2017.

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

[20] P. Vijayaraghavan and S. Vosoughi, "Tweetspin: Fine-grained propaganda detection in social media using multi-view representations," in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2022, pp. 3433–3448.

[21] A. M. U. D. Khanday, Q. R. Khan, and S. T. Rabani, "Detecting textual propaganda using machine learning techniques," *Baghdad Science Journal*, vol. 18, no. 1, pp. 0199–0199, 2021.

[22] M. Francisco, M. Á . Ben´ıtez-Castro, E. Hidalgo-Tenorio, and J. L. Castro, "A semi-supervised algorithm for detecting extremism propaganda diffusion on social media," *Pragmatics and Society*, vol. 13, no. 3, pp. 532–554, 2022.

[23] E. Ferrara, S. Cresci, and L. Luceri, "Misinformation, manipulation, and abuse on social media in the era of covid-19," *Journal of Computational Social Science*, vol. 3, pp. 271–277, 2020.

[24] H. Jeong, *Hierarchical Attention Networks for Fake News Detection*. The Florida State University, 2021.

[25] T. Zhang, D. Wang, H. Chen, Z. Zeng, W. Guo, C. Miao, and L. Cui, "Bdann: Bert-based domain adaptation neural network for multi-modal fake news detection," in *2020 international joint conference on neural networks (IJCNN)*. IEEE, 2020, pp. 1–8.

[26] K. Shu, S. Wang, and H. Liu, "Beyond news contents: The role of social context for fake news detection," in *Proceedings of the twelfth ACM international conference on web search and data mining*, 2019, pp. 312–320.

[27] H. T. Phan, N. T. Nguyen, and D. Hwang, "Fake news detection: A survey of graph neural network methods," *Applied Soft Computing*, vol. 139, p. 110235, 2023.

[28] S. Qian, J. Wang, J. Hu, Q. Fang, and C. Xu, "Hierarchical multi-modal contextual attention network for fake news detection," in *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, 2021, pp. 153–162.

[29] A. Thota, P. Tilak, S. Ahluwalia, and N. Lohia, "Fake news detection: a deep learning approach," *SMU Data Science Review*, vol. 1, no. 3, p. 10, 2018.

[30] C. Comito, L. Caroprese, and E. Zumpano, "Multimodal fake news detection on social media: a survey of deep learning techniques," *Social Network Analysis and Mining*, vol. 13, no. 1, p. 101, 2023.

[31] M. Abdul-Mageed, A. Elmadany, and E. M. B. Nagoudi, "Arbert & marbert: Deep bidirectional transformers for arabic," *arXiv preprint arXiv:2101.01785*, 2020.

[32] 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.

[33] I. Afyouni, Z. Al Aghbari, and R. A. Razack, "Multi-feature, multi-modal, and multi-source social event detection: A comprehensive survey," *Information Fusion*, vol. 79, pp. 279–308, 2022.

[34] M. N. Alenezi and Z. M. Alqenaei, "Machine learning in detecting covid-19 misinformation on twitter," *Future Internet*, vol. 13, no. 10, p. 244, 2021.

[35] M. G. Hussain, M. R. Hasan, M. Rahman, J. Protim, and S. Al Hasan, "Detection of bangla fake news using mnb and svm classifier," in *2020 International Conference on Computing, Electronics & Communications Engineering (iCCECE)*. IEEE, 2020, pp. 81–85.

[36] K. M. Yazdi, A. M. Yazdi, S. Khodayi, J. Hou, W. Zhou, and S. Saedy, "Improving fake news detection using k-means and support vector machine approaches," *International Journal of Electronics and Communication Engineering*, vol. 14, no. 2, pp. 38–42, 2020.

[37] A. M. Elmogy, U. Tariq, M. Ammar, and A. Ibrahim, "Fake reviews detection using supervised machine learning," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 1, 2021.

[38] A. Mahabub, "A robust technique of fake news detection using ensemble voting classifier and comparison with other classifiers," *SN Applied Sciences*, vol. 2, no. 4, p. 525, 2020.

[39] 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.

[40] R. Jehad and S. A. Yousif, "Fake news classification using random forest and decision tree (j48)," *Al-Nahrain Journal of Science*, vol. 23, no. 4, pp. 49–55, 2020.

[41] T. Dinesh and T. Rajendran, "Higher classification of fake political news using decision tree algorithm over naive bayes algorithm," *Revista Geintec-Gestao Inovacao E Tecnologias*, vol. 11, no. 2, pp. 1084–1096, 2021.

[42] N. L. S. R. Krishna and M. Adimoolam, "Fake news detection system using decision tree algorithm and compare textual property with support vector machine algorithm," in *2022 International Conference on Business Analytics for Technology and Security (ICBATS)*. IEEE, 2022, pp. 1–6.

[43] R. Gandhi, "Na¨ıve bayes classifier. towards data science," 2018.

[44] F. I. Adiba, T. Islam, M. S. Kaiser, M. Mahmud, and M. A. Rahman, "Effect of corpora on classification of fake news using naive bayes classifier," *International Journal of Automation, Artificial Intelligence and Machine Learning*, vol. 1, no. 1, pp. 80–92, 2020.

[45] N. S. Yuslee and N. A. S. Abdullah, "Fake news detection using naive bayes," in *2021 IEEE 11th International Conference on System Engineering and Technology (ICSET)*. IEEE, 2021, pp. 112–117.

[46] A. Yerlekar, P. Rohankar, A. Rakshit, P. Vairagade, S. Upase, and I. Gaharwar, "Fake news detection using machine learning approach multinomial naive bayes classifier," *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, vol. 9, 2019.

[47] S. Senhadji and R. A. San Ahmed, "Fake news detection using na¨ıve bayes and long short term memory algorithms," *IAES International Journal of Artificial Intelligence*, vol. 11, no. 2, p. 746, 2022.

[48] J. A. Nasir, O. S. Khan, and I. Varlamis, "Fake news detection: A hybrid cnn-rnn based deep learning approach," *International Journal of Information Management Data Insights*, vol. 1, no. 1, p. 100007, 2021.

[49] M. Umer, Z. Imtiaz, S. Ullah, A. Mehmood, G. S. Choi,

and B.-W. On, “Fake news stance detection using deep learning architecture (cnn-lstm),” *IEEE Access*, vol. 8, pp. 156 695–156 706, 2020.

[50] X. Zhi, L. Xue, W. Zhi, Z. Li, B. Zhao, Y. Wang, and Z. Shen, “Financial fake news detection with multi fact cnn-lstm model,” in *2021 IEEE 4th International Conference on Electronics Technology (ICET)*. IEEE, 2021, pp. 1338–1341.

[51] 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.

[52] T. A. Wotaifi and B. N. Dhannoon, “An effective hybrid deep neural network for arabic fake news detection,” *Baghdad Science Journal*, vol. 20, no. 4, pp. 1392–1392, 2023.

[53] H. Najadat, M. Tawalbeh, and R. Awawdeh, “Fake news detection for arabic headlines-articles news data using deep learning.” *International Journal of Electrical & Computer Engineering (2088-8708)*, vol. 12, no. 4, 2022.

[54] P. Hajek, A. Barushka, and M. Munk, “Fake consumer review detection using deep neural networks integrating word embeddings and emotion mining,” *Neural Computing and Applications*, vol. 32, no. 23, pp. 17 259–17 274, 2020.

[55] R. K. Kaliyar, A. Goswami, and P. Narang, “Echofaked: improving fake news detection in social media with an efficient deep neural network,” *Neural computing and applications*, vol. 33, pp. 8597–8613, 2021.

[56] N. Aslam, I. Ullah Khan, F. S. Alotaibi, L. A. Aldaej, and A. K. Aldubaikil, “Fake detect: A deep learning ensemble model for fake news detection,” *complexity*, vol. 2021, pp. 1–8, 2021.

[57] M. Berrahal, M. Boukabous, M. Yandouzi, M. Grari, and I. Idrissi, “Investigating the effectiveness of deep learning approaches for deep fake detection,” *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 6, pp. 3853–3860, 2023.

[58] Q. Li, Q. Hu, Y. Lu, Y. Yang, and J. Cheng, “Multi-level word features based on cnn for fake news detection in cultural communication,” *Personal and Ubiquitous Computing*, vol. 24, pp. 259–272, 2020.

[59] H. Saleh, A. Alharbi, and S. H. Alsamhi, “Opcnn-fake: Optimized convolutional neural network for fake news detection,” *IEEE Access*, vol. 9, pp. 129 471–129 489, 2021.

[60] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need. nips (2017),” *arXiv preprint arXiv:1706.03762*, vol. 10, p. S0140525X16001837, 2017.

[61] A. Gasparetto, M. Marcuzzo, A. Zangari, and A. Albarelli, “A survey on text classification algorithms: From text to predictions,” *Information*, vol. 13, no. 2, p. 83, 2022.

[62] 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.

[63] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever *et al.*, “Improving language understanding by generative pre-training,” 2018.

[64] C. Moyano Gravalos, “Deep learning on genomics using nlp-oriented algorithms,” Master’s thesis, Universitat Polite`cnica de Catalunya, 2023.

[65] T. S. Rao and C. Andrade, “The mmr vaccine and autism: Sensation, refutation, retraction, and fraud,” pp. 95–96, 2011.

[66] C. T. Mesquita, A. Oliveira, F. L. Seixas, and A. Paes, “Infodemia, fake news and medicine: science and the quest for truth,” pp. 203–205, 2020.

[67] E. Shushkevich and J. Cardiff, “Tudublin team at constraint@ aaai2021–covid19 fake news detection,” *arXiv preprint arXiv:2101.05701*, 2021.

[68] S. Shifath, M. F. Khan, and M. S. Islam, “A transformer based approach for fighting covid-19 fake news,” *arXiv preprint arXiv:2101.12027*, 2021.

[69] M. Samadi, M. Mousavian, and S. Momtazi, “Deep contextualized text representation and learning for fake news detection,” *Information processing & management*, vol. 58, no. 6, p. 102723, 2021.

[70] S. A. Yousif and R. Jehad, “Classification of covid-19 fake news using machine learning algorithms,” in *AIP Conference Proceedings*, vol. 2483, no. 1. AIP Publishing, 2022.

[71] F. Haouari, M. Hasanain, R. Suwaileh, and T. Elsayed, “Arcov-19: The first arabic covid-19 twitter dataset with propagation networks,” *arXiv preprint arXiv:2004.05861*, 2020.

[72] 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.

[73] P. Nakov, A. Barro´n-Ceden˜o, G. Da San Martino, F. Alam, J. M. Struß, T. Mandl, R. M´ıguez, T. Caselli, M. Kutlu, W. Zaghouani *et al.*, “The clef-2022 checkthat! lab on fighting the covid-19 infodemic and fake news detection,” in *European Conference on Information Retrieval*. Springer, 2022, pp. 416–428.

[74] H. Mubarak, S. Hassan, S. A. Chowdhury, and F. Alam, “Arcovidvac: Analyzing arabic tweets about covid-19 vaccination,” *arXiv preprint arXiv:2201.06496*, 2022.