Multimedia Tools and Applications (2024) 83:74469–74502
https://doi.org/10.1007/s11042-024-18564-7

Detecting fake news for COVID-19 using deep learning: a review

Hamza Zaheer[1] · Maryam Bashir[1]

Received: 5 April 2023 / Revised: 19 September 2023 / Accepted: 29 January 2024 /
Published online: 16 February 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract
The December of 2019, marked the start of one of the biggest pandemics that the human race had seen for some centuries. COVID-19 after originating from China was in full force and was spreading quickly. This, however, was different from the previous pandemics as this is the age of technology and social circles on the internet. Thus, a sinister form of situation arose where fake news and misinformation flooded social media. The situation got to the point that WHO termed it as an "infodemic". Thus, NLP was again implored to find a solution and massive research was conducted for the detection of fake news on these platforms. The success of fake news detection improved and by today i.e. in 2023 the techniques have matured quite a bit. Keeping both of these aspects in mind, we have conducted a detailed review on fake news detection techniques for COVID-19. We have discussed the collection of data by providing a deep analysis of 7 COVID-19 Fake News datasets. Moreover, during the analysis of different methodologies, domination of deep learning and hybrid models was observed - specifically ensemble of transformer based models. Additionally, we explored the practical implications of COVID-19 Fake News detectors as components in generative AI models and as browser extensions to keep the common people safe. Finally, we discussed the limitations in existing research and how it can be improved in the future by exploring multi-modal, feature rich and cross-lingual approaches.

Keywords Fake news · COVID-19 · BERT · Ensembles · Text classification

1 Introduction

COVID-19 is the new norm of the world and with this new norm, we must learn to live and thrive. The difference between this and previous pandemics such as black death or cholera is its longevity. They ceased to exist and yet COVID-19 has continued to make its impact on the world and is expected to be a part of it for generations to come. Therefore, seeing the effects that it has incurred, we must learn to live alongside them and in best case to overcome them.

✉ Maryam Bashir
maryam.bashir@nu.edu.pk

[1] FAST School of Computing, National University of Computer and Emerging Sciences, Lahore, Pakistan

One important aspect of COVID-19 was the emergence of fake news on social media sites. Fake news had been a feature of the internet for some time and was used extensively in election campaigns or to promote different views regarding a topic. However, during the COVID-19 pandemic the seriousness of the issue was really made apparent when the misinformation was classified as an "Infodemic" by WHO's Director General [1]. An extreme example of this was the spread of false news that alcohol cures COVID-19. This led to a 5 year old being poisoned due to an overdose of alcohol given as a cure by his parents [2]. While at other places, news was spread that 5g towers spread COVID-19 virus [3]. At one instance, this led to around 80 5g towers being destroyed by the public and many more overall. While working on the technology, many engineers were harassed and one had to be taken to the hospital due to being stabbed during work. The seriousness of the issue prompted the researchers to find an automated solution to the problem as manually curbing misinformation was proving to be too overwhelming.

Datasets for fake news were created and the community was in full flow. From conventional machine learning and deep learning models to innovative hybrid models were proposed and tested. Moreover, people explored new methods to enhance the performance of their models through clever data augmentation and representation techniques. In 2021, a shared task [4] was initiated to provoke the community to come up with more solutions to the misinformation problem. As the primary source of fake news was social media and internet, many browser extensions came to the fore claiming to flag false news. Finally, a research team, after much testing and deliberation, deployed a chrome extension of their model that would automatically detect fake news present on the page a user was visiting [5]. Furthermore, some thorough analysis was done on the nature of fake news and misinformation to understand it better and to help in the future research.

By far in fake news detection tasks, transformer models were widely used including state of the art BERT [6], RoBERTa [7] , XLNET [8] etc. followed by RNN based models such as LSTM, GRU along with their bidirectional variants. Moreover, extensive research on the machine learning models including Support Vector Machines, Logistic Regression, Naive Bayes and Decision Trees were also explored. Finally, the choice between different embeddings such as GloVe [9], Pre-Trained BERT embeddings and ELMO was also explored. The detailed survey of all the methods is provided in the rest of the paper.

1.1 Research objectives and contributions

The COVID-19 pandemic has had a profound impact on the world, and the large amount of information and opinions surrounding it has made fake news detection more important than ever. Fake news has caused mental health disorders in many people and compounded the suffering they were going through due to COVID-19. Many adverse health affects have resulted from misinformation and made the process of identifying a cure more difficult. A review on the topic can provide valuable insights and help advance the field. The objectives and contributions of this research are as follows:

1. **To describe fake news datasets:** this review presents details of different datasets for fake news detection which can help researchers to evaluate their techniques using these datasets.
2. **To identify gaps in the literature:** this review identifies limitations in the current research, including areas that have received limited attention and areas that are in need of further investigation.

3. **To facilitate comparison:** By presenting the results of multiple studies, this review facilitates comparison of different techniques for fake news detection.
4. **To guide future research:** By highlighting the strengths and weaknesses of existing research, this study points out areas that are in need of improvement and areas that offer the greatest potential for impact.
5. **To emphasize the uniqueness of COVID-19 Fake News:** By providing a comparative analysis of COVID-19 fake news and fake news in general, this study can highlight the unique challenges of the subject matter.

1.2 Disposition from the existing work

There have been some surveys conducted on fake news detection in past few years [10–14]. Among these surveys Zhang et al. [11] have done the most detailed and comprehensive survey on fake news. They present negative effects of fake news detection, the importance of fake news detection, details of different datasets, and features used in fake news detection. This paper was published in March 2020 and COVID-19 had just started at that time so they have not covered any study related to fake news detection for COVID-19. Zhou et al. [12] have also done a detailed study on fake news detection techniques but they have also not covered any COVID-19 studies as it was also published in 2020. Our survey is significantly different from these existing surveys as it is focused on fake news detection for COVID-19. COVID-19 has been around for almost four years now and there is need for detailed review on fake news detection related to COVID-19. COVID-19 pandemic has given people with ill intentions a great opportunity to spread all kinds of fake news about it. It is critical to detect these fake news in order to educate masses about true situation of the pandemic so that they can take necessary measures to prevent the spread. This review focuses on fake news studies related to COVID-19, present limitations in existing research, and future directions.

The rest of the paper is structured as follows: Section 2 describes the fake news, its impact on the world in COVID-19 and the unique nature of COVID-19 Fake News. Section 3 presents the details of datasets used in COVID-19 fake news detection, and methodologies are presented in Section 4. Next in Section 5 work done in languages other than English is discussed. In Section 6, we talk about the practical applications of research in COVID-19 fake news detection. In Section 7 and Section 8, we present limitations of existing work and highlight some of our suggestions in these fields and how they can be combined to form new studies. Finally, in Section 9 we provide our conclusion.

2 The plague of fake news in COVID-19

COVID-19 has continued to be a threat to the lives of many and is considered one of the worst pandemics that has hit the world in recent times. However, the pandemic has been made worse by the **infodemic** i.e. an abundance of information either fake or real. The term was made popular by the World Health Organization which has been at the forefront at fighting against the pandemic and now largely the infodemic (fake news related to COVID-19 pandemic). Fake news can be defined as: **"the news articles that are intentionally and verifiably false, and could mislead readers"** [15]. Fake news also includes any kind of false information on internet which is published intentionally to mislead the public [11]. Fake news is created mostly by humans but sometimes bots are also used to spread fake news through social media. Sometimes people without any bad intentions also participate in fake news spread by forwarding the fake information to more people without verification. The target of fake news are different groups of people depending on the agenda of the group or person spreading the news. The motivation for fake news is often political to gain votes, or to spread propaganda against certain religious groups etc.

Fake news related to COVID-19 started emerging in 2020 as the whole world was uncertain about the future and how long it will take to find a cure. Even when scientists were able to create vaccines for COVID-19, fake news related to adverse affects of the vaccination spread at a rapid rate. Many people in Europe refused to take the vaccine due to fake news that it can have long term non reversible adverse effects on health [16]. There was no evidence behind these rumors and people started to believe in these without any evidence. Some celebrities and famous people also contributed to the spread of misinformation as they constitute majority of social media engagement [16].

The infodemic has had many adverse effects on the mental awareness of people. According to a study by J. Torales [17], the people who are more exposed to COVID-19 news, whether real or fake, are 93% more prone to developing depressive disorders. Moreover, this is made worse by the lack of quality information on health organizations' sites as shown by [18]. However, the main culprit in all of the abundance of information are the social media sites. The abundance of self reporting and fake news on these sites have really made the worst of the worse. Therefore, these sites must start taking responsibility to curb fake news and maintain their platforms as demanded by [19]. On the other hand, another argument given by [20] is that people themselves should be trained on how to filter such kinds of news and individual responsibility must be enhanced so news is not shared without verifying the source and the information. This is augmented by a study done by [21] which shows that people with high cognitive ability look at the information quality while sharing the news whereas people on the other end of the spectrum rely on source authentication and quality of the argument. Moreover, fake news has impacted vaccine acceptance as well. This is shown in a study by [22] where people who accept fake news are more reluctant to be vaccinated. Moreover, this reluctance of vaccines has also impacted international travel. The type of vaccine determines to which countries one can travel as shown by [23].

Today COVID-19 might not be the beast that it was an year back, still the misinformation regarding it is spreading. Every coming day, a new variant is identified and we hear about people dying from it. Quarantine is still in effect at some places and people are suffering. Moreover, this is not the first time a pandemic has come and it will not be the last. Thus, research in this area would allow us to be ready for the next one. Moreover, models and research done on COVID-19 Fake News can easily be translated to other domains to detect fake news. All of these things prompt us to build an automated solution to COVID-19 fake news detection. People have started to understand the gravity of the situation and the research is catching up.

2.1 Unique challenges of fake news in COVID-19

Fake News has been a part and parcel of our everyday life since the inception of the internet or even before that. As long as the media has existed, fake news has as well. But what makes Fake News in COVID-19 different? Let's answer that question in this section.

- **Reach, Fear and Intensity:** The first factor is not about the challenges of detecting COVID-19 fake news, but about its potency. The three factors that make it so potent are its reach, fear and intensity. COVID-19 is a global phenomena, thus everyone knows about the challenges it poses. Moreover, everyone knows that COVID-19 can cause deaths which adds to its fear factor. Additionally, the entire world came to a halt when COVID-19 caused mass quarantines, this led to even more dissemination of Fake News and fear. Finally, all of these contributed to the intensity of the Fake News being spread. It was more potent, spread easily and increased havoc in the existing dystopia caused by COVID-19. This was unlike any type of fake news that was spread before this phenomena and had a much profound effect. Therefore, immediate research in the area was initiated to curb the effect of fake news in COVID-19.

- **Highly Specific Domain:** Another profound difference of COVID-19 fake news was its highly specific domain. COVID-19 deals with health and medicine with their specialized terms and fields. Traditional fake news deals with generalized concepts and general machine learning models can easily cope with them. This was also observed by Whitehouse et al. [24] in their methodology as the existing knowledge bases could not cope with the COVID-19 fake news dataset because of the existence of these specialized terms. Some of these terms are COVID-19, coronavirus, influenza, pneumonia etc. Thus, new models needed to be developed which could understand these terms and evaluate if the text consisted of fake news.
- **Rapidly changing information landscape:** Unlike general fake news which may deal with more stable topics, COVID-19 fake news is rapidly changing due to new claims, trends and rumors. A real time approach is needed for COVID-19 fake news detection which can keep up with the fast pace of misinformation spread.

Although the pandemic is over, the ongoing research related to COVID-19 is still important. The impact of COVID-19 is different for various regions of the world and the end of emergency status does not mean that the virus has disappeared from the world. Governments and policy makers are still dealing with the aftermath of the pandemic. Fake news about COVID-19, its spread or its treatments can have real public health consequences. This is especially true if new variants of the virus emerge or if regional and localized outbreaks occur. Detecting fake news and understanding how it spreads can be helpful in strategies for responding to similar emergencies, ensuring more effective communication and countering misinformation. While the pandemic is over, the possibility of new outbreaks or resurgences cannot be ruled out. In such cases, rapid spread of misinformation can worsen the situation and hinder containment efforts. Now, as we have an understanding of the problem and the unique challenges it poses, let's look at some of the datasets that authors have curated over the time to cope with this problem.

3 Data collection and the availability of datasets

Data sits at the core of all artificial intelligence tasks and without it there would be no concept of learning. Therefore, in listening to the community many COVID-19 misinformation datasets were released. Multiple datasets, often with the same method of collection were found in the literature. The data sources were bimodal in nature, with one being social media sites and the other being news websites. This brought good diversity in the type of data points with social media sites having shorter documents compared to news websites and articles.

Although there were many custom datasets made by authors to help them in their novel approaches, only few were used extensively by other authors in their research papers. Therefore, we will only be focusing on these datasets namely: Constraint [25], CoAID [26], CTF [27], ANTi-Vax [28], CovID [29], FibVid [30] and ReCOVery [31]. We found all of these datasets to be benchmark datasets with quite rich structure and thought put into them. Now, we will discuss each one of them in detail. The rest of this section will be structured in such a way that for each dataset, we will discuss the data collection and annotation methods and the data statistics.

3.1 Constraint

Constraint dataset [25] was released back in November 2020 and was then made part of a shared task [4] in the CONSTRAINT 2021 workshop co-located with AAAI 2021. The task was extremely popular due to the havoc that COVID'19 had wreaked. People were extremely interested in it and it won the best task of the year award. Continuing with this popularity most of the papers we reviewed used constraint dataset for their experiments and it was the most widely used dataset of English COVID-19 Fake News detection - see Fig. 9.

3.1.1 Data collection and annotation

Constraint dataset primarily deals with data extracted from social media sites. Two types of information are identified by the authors:

- Fake News
- Real News

Fake news is defined by the authors as any news that is proven to be unverifiable. Whereas real news is any news that is correct and can be verified as being correct from other sources. According to the authors the initial search is done by looking for keywords pertaining to the topic of COVID-19. Moreover, only text of the news is retained and all other information is ignored including images, videos and audio. Separate approaches were used to collect fake and real news.

For fake news, web was searched for social media content which contained news related to COVID-19. This news was checked against already verified documents and then labeled as fake if it contradicted the verifiable true information. The process was repeated for all the major social media giants including but not limited to Twitter, Facebook and Instagram posts. Moreover, apart from manual annotation, Boomlive, Politifact and Snopes were consulted to determine the correctness of a piece of information.

Collection of real news was straightforward as the news posted on verified health organization accounts is inherently true. The authors targeted accounts of organizations such as WHO and Centers for Disease to find examples of real news. Moreover, online media platforms and state accounts were also scrapped to find real news related to COVID-19.

3.1.2 Data statistics

After collecting the aforementioned data, the authors of the paper put news from both classes side by side to extract some useful insights. One important feature of the dataset is its lack of bias towards either class which is missing in other datasets. There were **5,600** Real news samples, and **5,100** fake news samples, making a total of **10,700** data samples (Fig. 1). Moreover, if we look at intra class differences, then it is observed that the length of fake news is shorter compared to real news. Looking at the top 10 important words from both classes in terms of frequency, it was also observed that the words in both classes matched, which highlights the difficulty of separating the data points for classification.

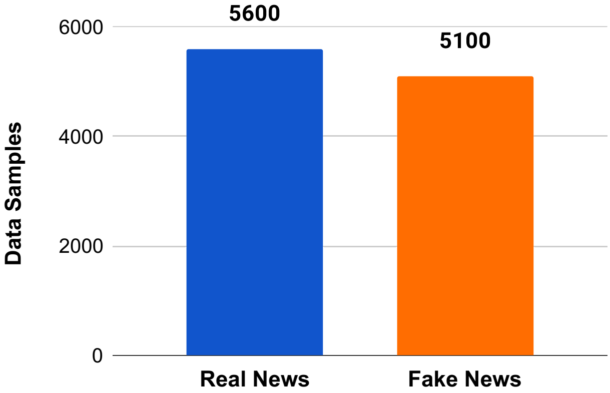

Fig. 1 Data distribution for Constraint dataset

3.1.3 Baseline results

The authors of the dataset also ran some baseline models on the dataset to get a feel for its nature. Four machine learning models were implemented: Gradient Boost, Support Vector Machine, Logistic Regression and Support Vector Machine with TF-IDF representation. Initial results showed that **Support Vector Machines** performed the best on the Constraint Dataset with an accuracy of **0.933** and can be used as a benchmark in future experiments.

3.2 CoAID

CoAID (COVID-19 he**A**lthcare m**I**sinformation **D**ataset) [26] is one of the relatively oldest COVID-19 related fake news dataset. It was released at the end of 2020 and marked the beginning of COVID-19 fake news detection. It uses data from December 2019 till September 2020 and has the ability to update its corpus automatically through the use of timestamps. This dataset, inspired by some other fake news datasets like Liar dataset [32], adopts their structure quite a bit as will be discussed later. It separates the data points into two categories:

- Claim
- News article

As the name suggests, claims refer to information that a person is claiming to be true and is made up of short sentences. On the other hand news articles are long pieces of writing that are written on a topic and provide detailed information.

3.2.1 Data collection and annotation

The data collection method used here is quite similar to Constraint dataset. For scraping, topics like COVID-19, flu, quarantine etc. are identified and shortlisted. Then news articles are collected from the web. For real news, reputable news brands are consulted and news reported by them is used as real news samples. However, for fake news articles, websites that provide facts are consulted. For collection of claims, WHO official accounts were scraped to get URLs of claims. After that, the content was retrieved using Newspaper3k [33].

Moreover, aside from the news articles and the claims provided by different organizations, another aspect of news was also explored i.e. self media. The argument provided was that the

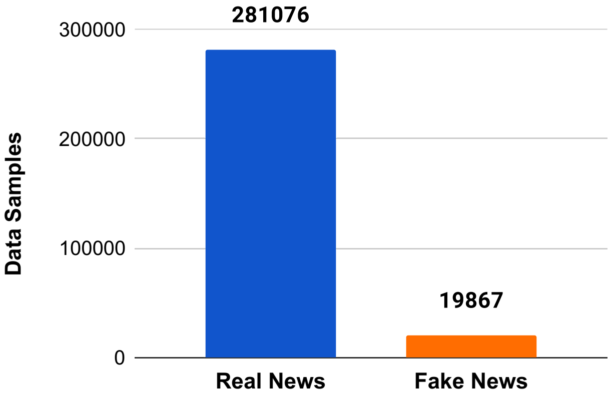

Fig. 2 Data distribution for CoAID dataset

users themselves also post and report their views on social media platforms. However, even this self reporting has a drastic impact in shaping the views of the general public. Therefore, tweets posted by individual users are also collected and added to the dataset.

After fetching the textual features using the above approach, user engagement features were also extracted. The motivation behind this approach was to collect metadata on top of the textual data so novel approaches could be developed using it. The user features include the tweets that were tweeted by users in response to the news about COVID-19. The authors took this one step further and also collected the replies to the tweets. Moreover, a mechanism for automatic updation of dataset was also introduced which uses timestamps to keep the versions organized.

3.2.2 Data statistics

As is evident from Fig. 2 and Table 1, the data collected is extremely biased towards true news. Moreover, sentiments of the collected corpus were also analyzed, and it was observed that the presence of negative sentiment was much more in fake news as compared to real news. This makes sense as the fake news is spread to cause anarchy and confusion in the public.

Table 1 Distribution of CoAID dataset

	Fake		Real	
	Claim	News Article	Claim	News Article
Information on Website	28	204	454	3,565
Tweets	484	10,439	8,092	141,652
Replies	626	7,436	12,451	114,820
Social Media Posts	650	–	42	–
Total	1,788	18,079	21,039	260,037

3.2.3 Baseline results

After collecting the dataset, the authors evaluated some models on the collected dataset. The selected models had a mix of machine learning and deep learning models which included SVM, CNN, HAN [34], dEFEND [35] etc. HAN [34] and dEFEND [35] use hierarchical attention networks. The best results were obtained using CNN, HAN [34] and dEFEND [35] with the best F1 score of **0.58** achieved using **dEFEND**.

3.3 CTF

CTF (COVID-19 Twitter Fake News) [27] was released in late 2021. This dataset is on the bulkier end of the spectrum and contains a lot of tweets for the community to use in their methods. The authors went into great detail to make a sizable dataset that could be used by state of the art deep learning techniques.

3.3.1 Data collection and annotation

The data collection and annotation was quite a rigorous process and took multiple steps to complete. In the first step the authors leveraged the already released COVID-19 tweets datasets. The released datasets were not labeled in the context of fake or real, however it did provide a large number of tweets to be annotated later. Moreover, twitter API was also used to extract tweets according to the relevant hashtags. Finally, there are datasets that are daily updated with COVID-19 tweets. They are used to collect the tweets as well. Moreover, data from official state departments and health authorities was also collected and labeled as being authentic.

After collecting the unlabeled tweets, labels were assigned to them. This is done by consulting fact checking websites like Politifact etc. These websites host a database of correct and false news. After scraping these websites, URLs are collected to build an annotated database.

In the third stage the database of annotated tweets is expanded. The intuition being that the false tweets refer to false information and the authentic tweets refer to correct information. This intuition is used to expand the labeled tweets by also considering the tweets that are referred to by the tweets collected in the previous stage. Moreover, in this step, two models are also trained on the data from the previous step to label the tweets collected in the first step. The models used were BERT [6] and RoBERTa [7] and were the major contributors in automating the entire process.

Finally, in the fourth step three human annotators are employed to check the labeled tweets in the previous step and to authentically annotate them. Three rules were used by the annotators for the entire process.

- A tweet is labeled as fake if it is against some known information that is correct.
- Or if it uses some known misinformation to support itself.
- Or if it gives the feeling of promoting misinformation by being humorous.

3.3.2 Data statistics

Some interesting analysis was performed on the collected dataset. It was observed that fake news highlights hashtags that are propaganda creating or are controversial in nature. On the other hand, genuine news highlights hashtags that are positive in nature and promote a positive thinking. Moreover, number of URLs were observed in the dataset and a conclusion was found that genuine news has a higher number of supporting URLs compared to fake news. Another, useful insight is the number of likes and dislikes. Genuine news had a considerably larger number of likes and retweets compared to fake news, highlighting the possibility of using this as a feature for classification. All in all **28,577** fake tweets and **28,790** real tweets were collected making it one of the largest COVID-19 fake news dataset. Moreover, the stats, Figure 3, shows its balanced nature as well.

3.4 ANTI-Vax

In 2021, vaccines for COVID-19 had finally been developed and governments all around the world had started a campaign to administer COVID-19 vaccines to their masses. However, with this came outcry from the public and the conspiracy theorists started floating made up news about the vaccines. This led to people rejecting the idea of vaccines and their benefit was being overlooked. Therefore, it was necessary to overcome all of the fake news surrounding vaccines. The first step against the fake news regarding vaccines was taken in the form of ANTI-Vax [28], a dataset tailored made for COVID-19 vaccines fake news. It was released in early 2022 and was one of the latest dataset released in this genre. The tweets included were tweeted during a span of 6 months that ended in July 2022. The authors followed a rigorous process to collect and annotate the tweets regarding the issue.

3.4.1 Data collection and annotation

For the data collection, some common myths regarding vaccines were collected first. These were taken from various credible sources, mainly from health organizations. After collecting these myths, criteria was defined to label a false tweet. A tweet was labeled false if it contained one of the myths and did not contain humor. Using this criteria many fake tweets were collected. The tweets that did not meet this criteria were labeled non misinformation.

The second step to the data verification step was the inclusion of domain experts. Many medical practitioners were consulted regarding the myths and some manual annotation was done to avoid any noise or errors in the collected dataset. This made the collected data more

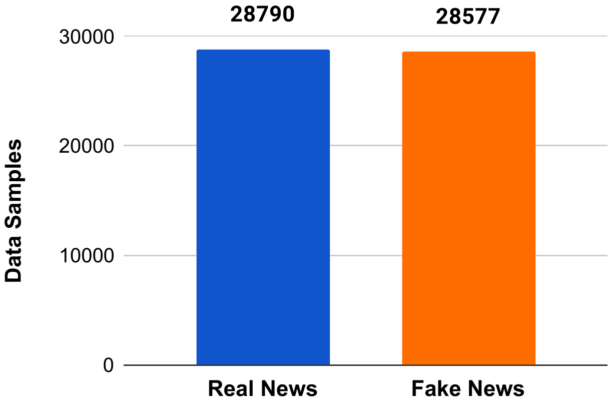

Fig. 3 Data distribution for CTF dataset

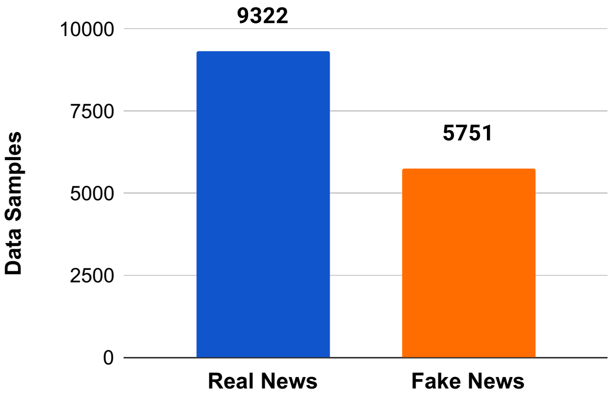

Fig. 4 Data distribution for ANTI-Vax dataset

robust and accurate. Finally, the number of fake news samples collected were **5,751** and the real news samples were **9,322** as shown in Fig. 4.

3.5 CovID

Fake news detection had been carried out on English textual data for some time. In the meantime, datasets for multilingual process were also coming on to the scene. Taking inspiration from them the authors of Coronavirus Infodemic Dataset (CovID) [29] decided to build a dataset around the multi modal nature of data. We all know that real data is rich with many elements, with humans using multiple means to present their ideas. These means include the written text, the visuals, the sound, the motion, and many more. However, the two most important among these are the textual data and the visuals. Therefore, the authors of CovID decided to use both of these facets of online data to build their multimodal fake news dataset. The dataset includes two version CovID I and CovID II. We will discuss both of them in the following sections.

3.5.1 Data collection and annotation

Before any machine learning and deep learning model is employed for fake news detection, the first step is to collect data. Therefore, special care must be taken to collect accurate data as any error here would be forward to the next steps and would ruin the entire proposed solution. For CovID, the authors have collected the data from three sources which are almost common to other datasets as well. The first source is “Poynter” [36], a fact checking website. Poynter houses a database of COVID-19 related news with already appointed labels to it. URLs of these news are extracted and then the content is used to build the dataset. The URLs belong to both news articles as well as social media posts from Facebook, Twitter etc. Moreover, both real and fake news is extracted using this method but the amount of real news collected using this method is limited.

Due to the limitation discussed above, a separate source was required for the extraction of real news. Therefore, official news websites were used as the second source of real news. Again keywords specific to COVID-19 were used to filter the retrieved news articles. A sizable amount of real news articles were retrieved in this manner. However, after some speculation another error was found in the approach. The fake news samples that were gathered had a good mix of news articles and social media posts. However, for real news only news articles were gathered. There is a measurable difference between the structure, writing, length, tone, audience and source of a news article compared to that of posts on social media. This poses a problem for the classifiers that would be trained on the dataset.

For optimal representation, some social media posts containing real news must also be added to the corpus. This brings us to the idea of CovID I and CovID II. The authors define CovID I as the dataset that contains samples till the previous step only i.e. no extra real news samples from social media sites are added. Whereas, CovID II incorporates social media posts as well. For social media posts, Twitter is consulted which is the largest social media platform. Twitter API is used to fetch the real news posts from the website. The tweets are collected from official accounts that are known to provide authentic and correct news. Once these tweets are extracted these are added to the corpus.

3.5.2 Data statistics

As is evident from Fig. 5, the authors have made an effort to make both of the datasets balanced with approximately equal number of samples for both of the classes. Moreover, it can be observed that real news articles in CovID II are greater in number compared to CovID I. This is expected, as CovID II contains additional tweets as well.

3.6 FibVid

In 2020, Kim et al. [30] collected a dataset for COVID-19 fake news for better understanding of the issues at hand. However, using their novel approach they also collected claims that were not COVID-19 related and built a combined dataset. Moreover, they not only collected the dataset, but did a deep analysis of the data to find patterns that are inherent in the COVID-19 fake news. Moreover, in addition to the collected fake news text, the authors also collected the user features as well as the features related to the spread of the news. We have discussed each of these aspects in the next sections.

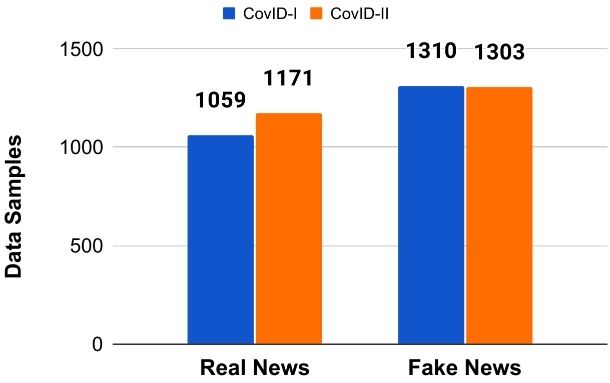

Fig. 5 Data distribution for CovID dataset

Table 2 Summary of FibVid dataset

	COVID		Non-COVID	
	Real	Fake	Real	Fake
News Claims	203	569	150	431
Twitter Claims	43	141	30	81
Tweets Count	27,296	133,374	12,793	47,790

3.6.1 Data collection and annotation

Kim et al. [30] followed a detailed process for the entire corpus collection. First of all they utilized "Politifact" and "Snopes" platforms for the collection of claims. After collecting the claims, keywords are extracted from them. The second step then initiates in which tweets containing the extracted keywords are collected. After collecting the tweets, the original tweets to which retweets were made are also collected and the process continues. Finally, the user information of the people who made the tweets are extracted to form the final feature set. Now, coming to the labeling, the labels for claims provided by the news checking websites are not binary. "Politifact" uses six classes whereas "Snopes" uses a humongous fourteen classes. To circumvent this problem, the authors use results of a previous research by Khan et al. [37] to convert these multiple classes into only two i.e. fake or real.

3.6.2 Data statistics

LDA (Latent Drichlet Allocation) was used to extract topics found in both real and fake news. However, after thorough analysis it was found that the topics discussed in both of these classes were almost similar. Moreover, depth (the number of retweets of the claim) and virality (the rate at which people spread the news) were observed. As expected, both volume and virality of fake news claims was higher compared to real news. Moreover, user behavior was also observed regarding the COVID and non-COVID news. The people who spread COVID-19 news were much more influential in the social media space and tended to be heavy users of the aforementioned technology. The distribution of claims in the dataset are provided in the Table 2 and Fig. 6.

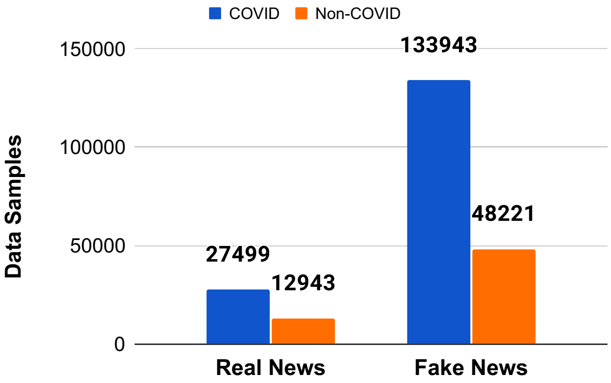

Fig. 6 Data distribution for FibVid dataset

3.7 ReCOVery

Like other datasets, ReCOVery was also released a few months after the huge wave of COVID-19 contractions. With the spread of virus, fake news was also spreading like wild fire. Thus, Zhou et al. [31] promptly got to work and released the dataset to advance research in the area. Taking extra care while preparing the dataset, they prioritized quantity over quality so, all the samples were correctly labeled. Moreover, they not only focused on textual features, but also advanced a multi-modal approach so unique features could be highlighted and used.

3.7.1 Data collection and annotation

Data annotation is a complicated process, often involving a lot of manual effort. However, Zhou et al. [31] came up with a clever idea to collect and annotate data while avoiding all of the manual effort. They relied on News Websites and vendors for their data collection and instead of annotating the news, they annotated the media outlets. To do this, they looked for media outlets on the extreme ends of the reliability scale. For this they relied upon NewsGuard [38] and Media Bias/Fact Check (MBFC) [39]. Both of these organizations do rigorous research and rate news outlets on a number of metrics to give them a final rating. By looking at their rating the authors skimmed through over 2,000 news outlets and finalized 60. Out of these 38 represented the fake news group and 22 the real news group.

After finalizing the news groups, the authors scrapped their articles and the reported news to look for COVID-19 keywords. The articles and reports which included the keywords were extracted. Along with the text of the articles, the cover images, authors, publisher, country etc. were also collected. After this process, the authors scrapped twitter to extract tweets mentioning these news articles. The tweets gave opportunity to the authors to add more features in the dataset by incorporating them into the dataset. The added features included tweet text, the user id of the person who tweeted, the network effect of the tweet etc. All of these things were combined to form a single dataset that could be used further for data classification.

3.7.2 Data statistics

While collecting the dataset, the authors uncovered some interesting statistics to help in further research. Most of the articles had less than 5 authors/collaborators and rarely did they find an article with an appreciable number of authors. Moreover, the number of news articles published exploded as the number of patients for COVID-19 were increasing. Also most of the news collected was concentrated in developed countries e.g. United States. Finally, the news articles collected were highly biased towards real news. With real news articles being **1,364** and fake news articles being **665**. The detailed data distribution can be seen in Fig. 7.

3.7.3 Baseline results

The authors also performed some experiments, to look at the baseline performance of some classifiers. Deep learning models were employed and they gave a decent performance on the dataset. In a group of LIWC+DT, RST+DT, Text-CNN and SAFE, SAFE gave the best results of **0.833** F-1 score on the Real News and **0.672** F-1 Score on Fake News. LIWC+DT was second with rest of them following the way.

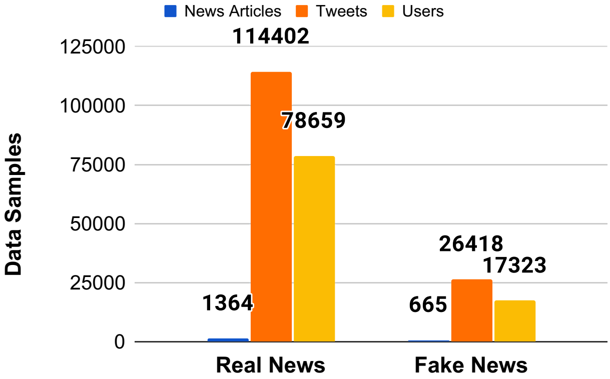

Fig. 7 Data distribution for ReCOVery dataset

As can be observed a detailed analysis of all 7 datasets were given. The Constraint dataset being the most popular one, followed by CoAID and then the rest follow - shown in Fig. 9. A basic data collection procedure of the datasets is provided in Fig. 8. These datasets form the go to resource for anyone who wants to research fake news detection in COVID-19. They are comprehensive yet they provide enough diversity to check for the generalization capabilities of the proposed algorithms for fake news detection. Finally, we have provided Table 3 to outline the dataset sizes and the distribution of samples within them (Fig. 9). Now, as the datasets are introduced, we will move towards the various approaches people have employed to tackle the problem. Section 4 will walk you through all the approaches discussed.

4 Methodologies

As the datasets for COVID-19 fake news detection started to become available, they were followed by a slew of people trying their approaches to tame and conquer this new problem. Thus, the amount of research done on this topic in the last three years has been amazing and we have seen quite a few novel approaches for solving this problem. Some people used traditional machine learning algorithms, discussed in Section 4.2, for the problem. While

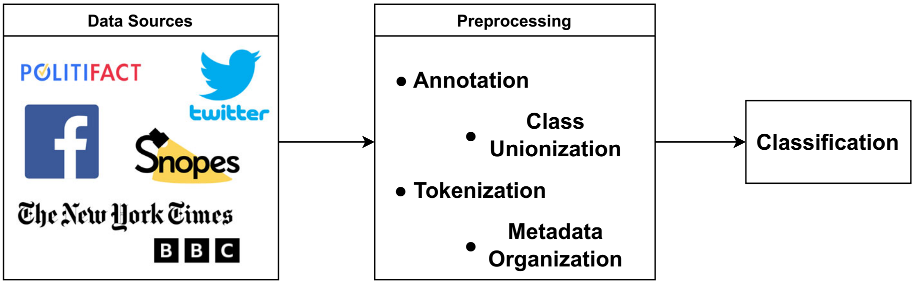

Fig. 8 Basic Data Collection Procedure

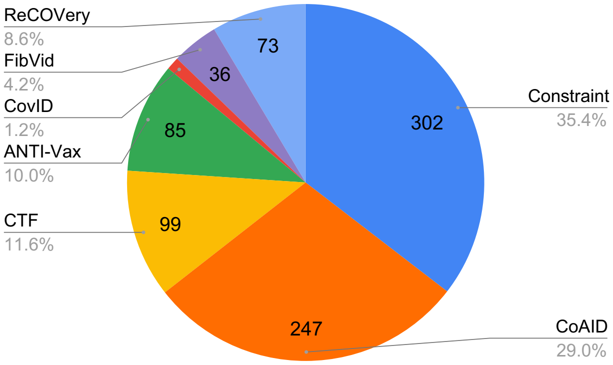

Fig. 9 Reference Count of Mentioned Datasets

a majority of them proposed deep learning approaches, discussed in Section 4.3. However, there were some people trying a mix of the two approaches as well, discussed in Section 4.4. We have tried to limit our search to the papers published in the years 2021 to 2023. This has

Table 3 Comparison of Distributions and Features of Different COVID-19 Datasets

Study	Dataset	Total Samples	Fake/Real News Ratio	Multi-Modal	User Features
Fighting an Infodemic: COVID-19 Fake News Dataset [25]	Constraint	10,700	0.477 / 0.523	No	No
CoAID: COVID-19 Healthcare Misinformation Dataset [26]	CoAID	300,943	0.067 / 0.933	No	Yes
Cross-SEAN: A Cross-Stitch Semi-Supervised Neural [27]	CTF	57,367	0.498 / 0.502	No	Yes
ANTi-Vax: a novel Twitter dataset for COVID-19 vaccine misinformation detection [28]	ANTI-Vax	15,073	0.382 / 0.618	No	No
ARCNN framework for multimodal infodemic detection [29]	CovID-I	2,369	0.553 / 0.447	Yes	No
ARCNN framework for multimodal infodemic detection [29]	CovID-II	2,474	0.527 / 0.473	Yes	No
FibVID: Comprehensive fake news diffusion dataset during the COVID-19 period [30]	FibVid	161,442	0.830 / 0.170	No	Yes
ReCOVery: A Multimodal Repository for COVID-19 News Credibility Research [31]	ReCOVery	2,029	0.328 / 0.672	Yes	Yes

the advantage that new methods are given more importance as they are the ones achieving state of the art results on fake news detection tasks. Finally, after discussing all of these approaches, we will be providing a brief analysis of the strategies discussed.

4.1 The necessity of data preprocessing

Before any method of text classification is used, the data must be preprocessed so it can be changed into a shape that the models can understand. Moreover, inherently data contains noise so by preprocessing it we are essentially removing all the noise that would have otherwise interfered with the classification task. The basic preprocessing steps that we found are listed below

- Tokenization
- Punctuation Removal
- Stopwords Removal
- Emojis Removal/Replacement
- Hashtags Removal/Replacement
- Mentions Removal

4.2 Traditional machine learning a viable option

In the last decade or so traditional machine learning has been shrugged to the side after the use of deep learning models in natural language processing (NLP) tasks. However, from our research they are still being used for NLP tasks albeit to determine baseline results only that are improved later by deep learning approaches. Some of the notable contributions in this regard are mentioned further. However, first let's talk about the data representation techniques used.

Initially, before training a machine learning model, we need to decide upon an effective **data representation** technique. In this regard Malhotra et al. [40] explored **T**erm **F**requency-**I**nverse **D**ocument **F**requency **(TF-IDF)** and Bag of Words **(BoW)** models on Constraint dataset [25]. While Bag of Words model represents each text by the frequency of tokens that occur in it, TF-IDF employs a more complex solution. In addition to the frequency of tokens, it also considers the inverse of document frequency - the inverse of text frequency. Moreover, Gundapu and Mamidi [41], took this a step further by using more complex word embeddings for Constraint Dataset. They used GloVe [9] vector embeddings, which make use of the co-occurence matrices to grasp the relationship between different tokens. As a twist, they multiplied the GloVe vector embedding of a token with its TF-IDF score to get the final representation of that token. For the experiments, they used **300** dimensional Glove embeddings. As increasing the number of dimensions has the effect of increasing the information stored as well.

With the hastle of data representation sorted out, Malhotra et al. [40] moved their sights towards the machine learning model to use. The machine learning algorithms considered in their study were:

- **S**tochastic **G**radient **D**escent **(SGD)**
- **B**ernoulli **N**aive **B**ayes **(BNB)**
- **K** **N**earest **N**eighbor **(KNN)**
- **D**ecision **T**rees **(DT)**
- **R**andom **F**orest **(RF)**

- **L**inear **S**upport **V**ector **M**achines **(LSVM)**
- **L**ogistic **R**egression **(LR)**

Each of these models were trained and then tested on Constraint dataset. For the majority of classifiers TF-IDF performed better as it is a better representation of informative words. In addition to these, some deep learning models such as **R**ecurrent **N**eural **N**etwork **(RNN)**, **L**ong **S**hort **T**erm **M**emory **(LSTM)** and **G**ated **R**ecurrent **U**nit **(GRU)** were also implemented. However, LSVM showed its supremacy with an accuracy of **0.941** being achieved using the TF-IDF representation. This was expected as historically **S**upport **V**ector **M**achines **(SVMs)** have performed better on NLP tasks. Gundapu and Mamidi [41], confirmed these findings on their own experiments on Constraint dataset. Along with their novel data representation technique using GloVe and TF-IDF, they used SVM, **P**assive **A**ggressive **(PA)** Classifier and **M**ulti **L**ayer **P**erceptron **(MLP)**. Again, SVMs gave the best accuracy of **0.964** which is a difference of around 0.02 with the best achieved by SVM using other embeddings. The supremacy of SVMs is again confirmed by Thaipisut et. al. [42], as they too achieved best results using SVMs. In addition to the Constraint dataset they also ran the machine learning algorithms on a Thai language dataset and on both of them SVMs gave the best result.

Another approach was used by Mehta and Mishra [43], to apply machine learning models on a custom dataset. They made their dataset using fact checking websites and ran **G**radient **B**oosting **(GB)**, LR, RF, and DT using TF-IDF sentence representation. Best results were achieved using LR with an F1 score of **0.930** on fake news.

In a different kind of study, Mazzeo et al. [44] wanted to investigate the effect of URL features on the classification accuracy. Not content with existing datasets, they devised a custom dataset by querying web search engines. In addition to news samples, metadata features were gathered these include domain features, word features, host features and lexical features. These features were fed into a feature selection algorithm i.e. **Chi-Squared** and **Pearson's Correlation Coeffecient** to choose the most appropriate features. Finally, the features were fed into various machine learning models with **N**aive **B**ayes **(NB)** getting the highest F-1 score of **0.810** on the dataset. Quite similar to this approach, Ahmad et al. [45] used evolutionary algorithms for COVID-19 fake news detection. Again they used a rather novel dataset for fake news detection released by Koirala [46] for their experiments. By using different types of stemmers and data representation techniques like TF-IDF and BOW, they created 6 different representations of news. Finally, they used KNN as a classifier and three meta heuristic based feature selection techniques i.e. **B**inary **G**enetic **A**lgorithm **(BGA)**, **B**inary **P**article **S**warm **A**lgorithm **BPSA** and **B**inary **S**alp **S**warm **A**lgorithm **(BSSA)**. KNN achieved the best accuracy of **0.754** using BGA with TF-IDF features and Lovins stemmer [47].

These were all the traditional Machine Learning Algorithms that provide quite respectable results in the task of COVID-19 Fake News detection - see Table 4 for a comprehensive overview. The following section will talk about the deep learning methods.

4.3 The big guns of deep learning

Although traditional machine learning techniques provide acceptable results for most of COVID-19 fake news detection tasks, they are still not the best. Being a sub-set of Machine Learning Algorithms, Deep Learning revolutionized the field of Natural Language Processing (NLP). The rule based approaches were over and the domain started moving towards the automated processes of deep learning. Machine translation, spam filtering, text generation and many other tasks saw great improvements through employing deep learning techniques.

Table 4 Traditional Machine Learning Algorithms and the Datasets Used

Ref.	Dataset	Algorithm	Accuracy	F1-Score	AUPRC	AUROC
Gundapu and Mamidi [41]	Constraint	LSVM	0.940	0.940	–	–
Patwa et al. [25]	Constraint	SVM	–	0.933	–	–
Mehta and Mishra [43]	Constraint	LR	–	0.930	–	–
Mahlous and Al-Laith [48]	Custom Arabic	LR	–	0.933	–	–
Wang et al. [49]	Custom Chinese	Random Forrest, SVM	–	–	0.230 (Random Forrest)	0.770 (SVM)
Mazzeo et al. [44]	Custom English	Naive Bayes	–	0.810	–	–
Shushkevich et al. [50]	Custom English	RF	0.790	–	–	–
Al-Ahmad et al. [45]	Koirala	KNN-BGA	0.754	–	–	–

From RNNs, LSTMs and GRUs to state of the art transformer models, deep learning has continued to be the go-to technique to tackle all language tasks. Therefore, it was natural for researchers to employ these techniques in the sensitive issue of COVID-19 fake news detection. In this section, we will aim to provide a walk-through into the techniques used by researchers to construct solutions around deep learning.

The first approach used in the literature was to try baseline deep learning approaches for fake news detection. This was achieved by Wani et al. [51] where they chose the Constraint dataset for all their experiments. The models for the initial testing were

- **C**onvolutional **N**eural **N**etwork **(CNN)** – Based on the mathematical operation of Convolution
- **L**ong **S**hort **T**erm **M**emory Network **LSTM** – an advanced form of Recurrent Neural Networks
- Bi-LSTM with Attention – A bidirectional version of LSTM with the attention mechanism which allows the model to focus on different parts of the input sequence
- **H**ierarchical **A**ttention **N**etworks **(HAN)** – Based on LSTM with multi layers including word and sentence encoder accompanied by corresponding word and sentence level attention
- **B**idirectional **E**ncoder **R**epresentations from **T**ransformers **(BERT)** – A transformer model that is tailored built with attention in mind and uses multiple linear layers along with attention layers
- DistilBERT – An alternative to the BERT model, it is simpler in nature with way less layers and thus less parameters

All of these models were run on the dataset and results were gathered. However, as the transformer based models are pre-trained on a large corpus of text therefore, to add another step they were trained on the unlabeled COVID-19 fake news as language models. Apparently, this was quite beneficial as the best results were obtained by BERT after it had been trained as a language model. The accuracy achieved was **0.984**. It was followed by DistilBERT with an accuracy of 0.982. However, rest of the models could only achieve an accuracy of around 0.940.

Hande et al. [56] too ran baseline results on the Constraint dataset - albeit all the models used were based on transformers. The authors also wanted to know the effect of these transformer based models when pre-trained on COVID-19 related news rather than language pertaining to general topics. They used BERT, DistilBERT, ALBERT, RoBERTa, XLNET, BART, ELECTRA and **C**OVID-**T**witter BERT **(CT-BERT)** as the topic specific transformer model. Moreover, in addition to these models the authors introduced a new loss function for this problem. Traditionally, for binary classification tasks **B**inary **C**ross **E**ntropy **(BCE)** Loss is used. On the other hand, if the classes are very unbalanced then Dice Loss is also used. However, the authors have combined both of these losses to form a custom **BCE-Dice** Loss. This had a positive effect on the results and after running the experiments CT-BERT along with the custom loss function were the best performing, bagging an accuracy of **0.982**. Which is a difference of around 0.01 compared to other models. This proves that topic specific transformer models are better performing than the rest.

As is evident from the above discussion, topic transformers are better. Therefore, Whitehouse et al. [24] decided to take this a notch further. Instead of taking models that were pre-trained on topic specific corpus, they sought out methods to incorporate relevant knowledge bases into existing models. This has the effect of making the models robust against adversarial methodologies. As they adapt their fake news to deter existing models, new knowledge bases can be added to detect this changed fake news as well. Second effect is that the model learns to compare new news with knowledge bases and if a difference is found, the identification of fake news would be easier. Keeping this in mind, the authors went on to incorporate knowledge bases using four models ERNIE, KnowBert, KEPLER, K-ADAPTER. They used the LIAR dataset [32] for non-COVID-19 news and the Constraint dataset for news related to COVID-19. The experiments showed improvements on LIAR dataset, in terms of accuracy, but the results were not too promising on the Constraint dataset. This was attributed to the stylistic differences between the Constraint dataset and the knowledge bases used. Moreover, Constraint dataset has features like the number of links which can conclusively determine the presence of fake news and the effect of knowledge bases is reduced. Among the models K-ADAPTER was the best performing with an accuracy of **0.981** on the Constraint dataset and KnowBert had the best accuracy of **0.2895** on the LIAR dataset.

Now, moving away from the transformer models a bit, Karnyoto et al. [58] proposed a method based on counterfeit news generation for augmenting the dataset. They used **L**atent **D**irichlet **A**llocation **(LDA)** for topic selection and then used BART as a counterfeit generator for the topics. Then cosine similarity was used to check the similarity of the generated text with the original text. If it was below the threshold of 0.950 then the counterfeit example is added to the dataset. Finally, CNN and LSTM were trained on the Constraint dataset to check the results. Both CNN and LSTM increased their performance with LSTM achieving an F1-score of **0.961** on the fake samples. Another similar approach involving LDA was used by Gautam and Masud [57]. They proposed a method to combine topic distribution of LDA with embeddings created by XLNET. The proposed model performed better compared to other models and achieved an F1-score of **0.967** on the Constraint dataset. Karnyoto et al. [59] had another go at the Constraint dataset with another augmentation technique. They augmented features of BiGRU-CRF with those of BiGRU-Att_CapsuleNet and used BERT and GPT2. Although the accuracy of **0.9196** was better compared to the baseline results, it was not good enough for the Constraint dataset. Yet another approach used by Karnyoto et al. [55] was the use of graph neural networks and simple data augmentation on the Constraint dataset. The proposed method used random deletion, random insertion, random swapping and synonym replacement for the data augmentation part. For the graph neural network, word to word and document to word edges were proposed. The models used were Graph Convolutional Networks, Graph Attention Networks and GraphSAGE. The best result of **0.928** was obtained using GraphSAGE which is better than LSTM and CNN but still not close to the capability exhibited by transformers.

Apart from Constraint dataset, other datasets were also researched by Heidari et al. [53]. They explored the addition of bot features into the feature set of fake news datasets. It is known that bots correspond to a huge degree of fake news dissemination and thus was used as the core idea of the paper. Features like bot or not bot are added along with fake claim scores. Moreover, metadata such as followers count, retweet count etc. were also added. Finally experiments were performed on the Cresci 2017 [68] dataset and the CoAID dataset. The results showed that the addition of bot features did not contribute much but the fake claim score increased the accuracy by 0.03. Another such approach was used on the CTF dataset by W. S. Paka [27] where textual features were augmented by user features and external knowledge for the classification task. Moreover, a cross stitch method was used to mix and match these features. The resulting model achieved an accuracy of **0.950** which was a difference of more than 0.09 compared to other models.

It is well known, that deep learning models are hard to explain and work like a black box. However, a number of authors went out of their way to build **explainable architectures** that could be understood easily. The first take on this was the approach used by Vijjali et al. [62] to

Table 5 Deep Learning Algorithms and the Datasets Used

Ref.	Dataset	Algorithm	Accuracy	F1-Score
Hayawi et al. [28]	ANTi-Vax	BERT	–	0.980
Ameur and Aliane [52]	AraCOVID19-MFH	ARABERT Cov19	–	0.958
Heidari et al. [53]	CoAID	BERT	0.860	–
Cui and Lee [26]	CoAID	dEFEND	–	0.581
Kou et al. [54]	Constraint, CoAID	HC-COVID	0.939 (Constraint), 0.899 (CoAID)	0.938 (Constraint), 0.820 (CoAID)
Karnyoto et al. [55]	Constraint	GRAPHSAGE	–	0.928
Wani et al. [51]	Constraint	BERT	0.984	–
Hande et al. [56]	Constraint	BERT + Dense + BCE Dice Loss	0.982	–
Gautam et al. [57]	Constraint	XLNet + Topic Distributions(Using LDA)	–	0.967
Karnyoto et al. [58]	Constraint	LSTM	–	0.961
Karnyoto et al. [59]	Constraint	BERT-BiGRU-Attention-CapsNET-(BiGRU-CRF)	0.9196	–
Thaipisutikul et al. [42]	Constraint, Custom Thai	Transformer	0.952 (Constraint), 0.940 (Custom Thai)	–
Gupta et al. [60]	Constraint, Hindi Hostility	SVM, Hindi BERT Embedding with BERT	–	0.935 (Constraint), 0.970 (Hindi Hostility)
Whitehouse et al. [24]	Constraint, LIAR	K-ADAPTER +. RoBERTa	0.981 (Constraint), 0.2895 (LIAR)	–
Bang et al. [61]	Constraint, Tweets-19	RoBERTa	–	0.981 (Constraint), 0.543 (Tweets-19)
Vijjali et al. [62]	Covid-19 Claims	BERT+ALBERT	0.855	–
Cheng et al. [63]	Covid-19 Rumor	BERT + LSTM VAE	–	0.8598
Paka et al. [27]	CTF	Cross SEAN(BERT and attention)	–	0.95
Sarnovskỳ et al. [64]	Custom Slovak	CNN+biLSTM	0.989	–
Mookdarsanit and Mookdarsanit [65]	Custom Thai	ULMFiT	–	0.729
Mattern et al. [66]	FANG-COVID	BERT + Social Context	0.824	0.683
Zhou et al. [31]	ReCOVery	SAFE	–	0.672 (Fake News), 0.833 (Real News)

Table 5 continued

Ref.	Dataset	Algorithm	Accuracy	F1-Score
Du et al. [67]	ReCOVery, Fake-Covid, CoAID used for training, Custom chinese for testing	CrossFake(BERT based)	–	0.735

detect COVID-19 fake news. The method is based on finding the entailment of a text from an explanation. Two transformer based models are trained on a custom dataset which contains news and its corresponding explanation. Model A fetches the most important explanations for a claim which are filtered using mean and standard variance thresholds. Model B checks the entailment of the claim from the explanation provided by model A. If the claim can be inferred from the explanation then it is a valid claim otherwise it isn't. Using this strategy, a number of transformer based models were trained on the COVID-19 Claims dataset, also introduced by the authors. However, a combination of BERT and ALBERT achieved the highest accuracy of **0.855** on the test set. A similar approach was used by Magistris et al. [69] where they put a claim through a number of steps to find its category and then to find its sub category. After this, documents of the determined sub-category are fetched out of which top n documents based on the similarity with the news are shortlisted. Then it is determined if the n documents agree with the news or disagree. If most of them agree then the news is valid otherwise it is not valid. Another similar approach was by Kou et al. [54] where the HC-COVID framework is proposed, a four stage architecture based on Graphs and Attention mechanism. The framework takes claims to rebut or verify news from expert or non-expert reviewers. These claims are then used to automatically classify the fake news samples. To test the framework, experiments were performed on the CoAID and Constraint datasets. The HC-COVID framework achieved an accuracy of **0.899** and **0.939** on each of the datasets respectively.

For a comprehensive view of the strategies discussed, see Table 5. Now, we have evaluated both traditional machine learning and deep learning approaches. However, in next section we will talk about the work put into combining multiple approaches to form a single hybrid algorithm to solve the problem.

4.4 The goods of many in hybrid models

Ensemble methods use prediction values for multiple models to make its final prediction. This incorporation of multiple models has the effect of minimizing the error of each of the models in making the final prediction. Thus, in essence, error of prediction and noise is reduced, as one's weakness is covered by others' strength. The same approach is used to tame the COVID-19 Fake News detection task and to achieve the highest results on it.

Coming back to the Constraint dataset, Biradar et al. [70] investigated multiple fusion models for COVID-19 fake news detection. In the first model, BERT, ELMO and XLNET were used to extract features from the text and finally their features were concatenated to form a final feature vector which is fed to multiple dense layers for classification. In the second model, LSTM, BILSTM, GRU and BIGRU were used to classify a text into fake or real classes. For the final prediction, average of their prediction values is taken. In the third model voting architecture is used. BERT, ULMFIT and a conventional classifier such as LR are used. The majority vote is taken as the prediction value. Lastly, bitwise OR of the prediction values of LR, BERT, SVM and KNN are taken to report the final prediction in the fourth model. All ensemble models were run but the best accuracy of **0.980** was achieved by the ensemble of LR, BERT and ULMFIT using the voting architecture.

Another very high performing model that utilizes voting architecture on the Constraint dataset was proposed by Glazkova et al. [71]. They used three CT-BERT models with different data splits and seeds. In the final stage they used hard voting to reach the final prediction. This ensemble of models gave them an excellent F-1 score of **0.987** on the Constraint dataset. Das et al. [72] also explored different types of voting architectures to power their ensembles. The ensemble models used multiple models to create multiple prediction vectors. The final prediction is then determined by either soft voting or hard voting. In soft voting the average value of the predictions given is used for the final prediction whereas in hard voting the majority vote is used as the final prediction. Moreover, aside from this ensemble model a couple of heuristics were also used. A conditional probability is calculated of the username of the user who posted the fake news to be malicious. For the second heuristic the same approach is used but for the domain of the news URL. The combined approach achieved state of the art results of **0.988** F-1 score. The individual models in the ensemble were RoBERTa, XLM-RoBERTa, XLNET and DeBERTa with hard voting architecture.

In addition to trying many traditional machine learning models, Gundapu and Mamidi [41] also proposed an ensemble of BERT, ALBERT and XLNET. The softmax values from these models were averaged to get the final prediction value. It was shown that this ensemble of models produced better results compared to individual models on the Constraint dataset by achieving an F-1 score of **0.986**. Last model to be used in the Constraint dataset was by Malla and Alphonse [73]. They chose the best performing individual models on the Constraint dataset i.e. RoBERTa and CT-BERT which have domain knowledge. The prediction vectors obtained by both models are multiplied together to form a matrix. Then the final prediction values are calculated from this matrix. The F-1 score of **0.989** achieved using this approach was the best that we found in the literature for Constraint dataset.

Aside from all these unimodal approaches, Raj et al. [29] proposed a multimodal approach for fake news detection. They used CovID dataset for their experiments along with some other benchmark COVID-19 fake news datasets like CoAID. They experimented with various model fusion approaches including early fusion, average fusion, sum fusion, max fusion and weighted average fusion for their image and textual models. At the end of the experiments, weighted average fusion method was the best performing. It calculates weights according to the performance of individual models. For visual features MobileNet V2 was the best and for textual features BiLSTM gave the best results. Combined, they achieved an accuracy of **0.929** on the CovID dataset. Another similar study was done by Gonwirat et al. [74] using the Koirala [46] and ISOT [75] datasets. They too proposed a weighted average method for the ensemble model of CNN-LSTM with various optimization functions. Using their approach they achieved an accuracy of **0.996** on ISOT and **0.737** on the Koirala dataset. For a comprehensive overview see Table 6. Figure 10 provides reference count of each model category.

5 The domain of other languages

COVID-19 fake news detection has extensively been done on English language and rightly so as it is the most widely used language in the online circle. However, focus must be given

Table 6 Hybrid Algorithms and the Datasets Used

Ref.	Dataset	Algorithm	Accuracy	F1-Score
Das et al. [72]	Constraint	RoBERTa + XLM-RoBERTa + XLNet + DeBERTa	–	0.988
Biradar et al. [70]	Constraint	LR + ULMFit + BERT	0.980	–
Glazkova et al. [71]	Constraint	Ensemble of CT-BERT	–	0.987
Malla and Alphonse [73]	Constraint	RoBERTa+CT-BERT	0.989	0.989
Gundapu and Mamidi [41]	Constraint	BERT + ALBERT + XLNET	–	0.986
Raj and Meel [29]	CovID	ARCNN(Bi-LSTM + MobileNetV2)	0.929	0.899
Gonwirat et al. [74]	ISOT, Koirala	CNN-LSTM	0.996 (ISOT), 0.737 (KOIRALA)	–

to other languages as well. Therefore, researchers from these local communities have strived to develop fake news detection techniques for their local languages.

For Hindi Devnagri script, Gupta et al. [60] proposed a hybrid model for hostility detection with a subcategory of it being fake. The models used were Hindi BERT, Hindi FastText and some meta data features which were combined to get an F1-score of **0.970**. Moving towards the Thai language, Mookdaranit and Mookdarsanit [65] collected Thai news samples for the fake news detection task. They investigated feature shifting along with pre-trained language models for text classification. To prepare the BERT [6], GPT [76] and ULMFIT [77] classifiers for Thai language, English to Thai translation was done of open English COVID-19 datasets. This translated data was used for pre-training of the models after which they were used for news classification. Another multilingual study on Moroccan tweets was undertaken by Madani et al. [78]. They tried multiple machine learning and deep learning techniques on their Moroccan tweets dataset extracted from Twitter. Moreover, they proposed that by adding

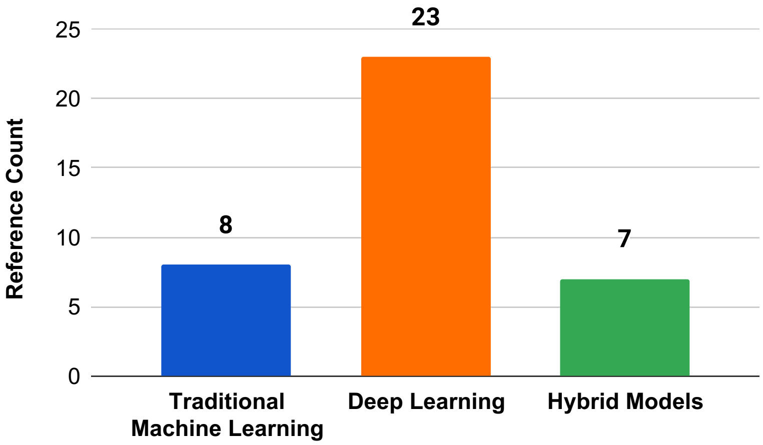

Fig. 10 Reference Count of Each Model Category

sentiment features into the feature space, better classification accuracy could be achieved. Their hypothesis turned out to be true and their model based on Random Forest Classifier beat traditional models for Moroccan fake tweets classification.

Arabic is one of the most widely spoken languages in the world. So, naturally a study was done by Mahlous et al. [48]. They collected a dataset of over seven million tweets in Arabic language. Fact checking websites were used for the annotation of the collected tweets. Moreover, a small handful of tweets were also manually annotated to be used as a baseline. After doing experiments with various Machine Learning algorithms and various representations, Logistic Regression with TF-IDF representation gave the best result of **0.878** F1-score on the manually annotated corpus. For automatically annotated corpus, it gave an even better score of **0.933** using bag of words representation. Another study involving Arabic language was undertaken by Ameur et al. [52]. They released a dataset named "AraCOVID 19MFH". It contained a total of 10,828 tweets in addition to some metadata attached to tweets as well e.g. positive/negative, factual, and the check worthiness etc. Moreover, the dataset had 10 labels corresponding to different aspects of the tweets which can be used for other tasks as well. Furthermore, they also fine tuned transformer based Arabic models such as AraBERT and mBERT for the COVID-19 dataset resulting in the models AraBERT-Cov 19 and mBERT-Cov 19. The resulting models gave the best weighted F-1 score of **0.958** on the fake news detection task.

A similar approach was used by Wang et al. [49] for the Cantonese language. An annotated dataset was collected using online discussion forums in Hong Kong and performance of various machine and deep learning approaches were investigated. Deep learning approaches once again proved to be better for the task. Following Cantonese, a study was done for Chinese language by Du et al. [67]. Their argument was that as most of the fact checking data is written in English therefore a cross lingual model is required. A BERT based classifier was trained on the English language. Then the Chinese language sentences were fed into the classifier after translation into English using Google API. Through experiments the authors formed a hypothesis that this approach was better than models which encode Chinese language directly which was proven right.

For German language, Mattern et al. [66] released a German dataset for COVID-19 fake news detection. The dataset contained a healthy number of around 41,000 tweets in German along with some other features as well, user engagement etc. The dataset was released with the intention of increasing research in German language for the aforementioned task. Finally, some baseline models were run on the dataset and the relevance of the collected social context features were checked. BERT along with the social context features was the best performing giving an accuracy of **0.981**. Sarnovskýet et al. [64] followed the approach for Slovak language. A Slovak language dataset was collected from the local news outlets and annotation was done. Finally, deep learning approaches were used for the classification. BiLSTM along with a convolution layer reached an F-1 score of **0.940**.

6 A practical aspect of COVID-19 fake news detection

Research in a field reaches new heights if it is backed up by strong practical applications. Same is the case with COVID-19 Fake News detection. It has a lot of practical implications that can be explored to advocate its research.

- **Removing false news from generative AI:** With the advent of generative AI like Chat-GPT[1], based on GPT-4 [79], and DALL-E 2[2], people have relied more and more on these models to get information. From getting steps to a cooking recipe to finding magical cure to COVID-19, all things can be queried on Chat-GPT. Keeping this in mind generative AI models need to be robust against fake news that could disturb the answers they produce. If you feed it garbage then it will spit out garbage. Therefore, fake news detectors can help in filtering out the fake news before it is fed into the generative models for training. Similarly, due to unknown reasons generative AI can produce false results as well which can be flagged during training to make it more robust to it. In the same manner, after training when model is deployed, we can have a fake news detector component which filters fake news from the models output. Thus, the fake news detector can help before, during and after training phase of the model. All of these can help in curbing fake news so, no user is harmed. A flow chart described in Fig. 11 can help in understanding the process.
- **A Fake News Detector Browser Extension:** In this era of social and mass media, wherever we go Fake News seems to follow us around. Thus, social media sites must take responsibility to curb the spread of fake news, especially related to COVID-19 as it deals with people's health. However, instead of putting responsibility on these giants and doing nothing, an individual must also take steps to keep away from such things. In this regard, a browser extension to detect, filter and report fake news can be extremely useful. An initiative was taken by W. S. Paka [27], in which they released a chrome extension which used their proposed model. Similarly, more extensions can be deployed to give the control to the user so, they can keep themselves safe from the unwanted anxiety caused by COVID-19 Fake News.

7 Limitations in existing research

There are several research gaps in fake news detection on COVID-19:

- **Limited cross-lingual support:** Most fake news detection algorithms are only designed to work for English language, leading to gaps in coverage of non-English speaking countries. This means that people who use social media or consume news in their native languages are missing out on a lot of this research.
- **Bias in training data:** The accuracy of fake news detection algorithms can be heavily influenced by the bias present in the training data. This is especially relevant in the context of COVID-19, where there is a large amount of conflicting information and opinions. Moreover, the datasets are also somewhat imbalanced and the experiments done might yield good weighted F-1 scores but that of Fake class are not that good.
- **Difficulty in identifying newly emerging fake news:** The fast-changing nature of the COVID-19 pandemic means that new forms of fake news are constantly emerging, and existing algorithms may struggle to keep up.
- **Integration with other sources of information:** In order to effectively detect fake news about COVID-19, it is necessary to integrate various sources of information, including official reports, media articles, and social media posts. This is a complex and ongoing challenge.

[1] https://openai.com/blog/chatgpt

[2] https://openai.com/dall-e-2

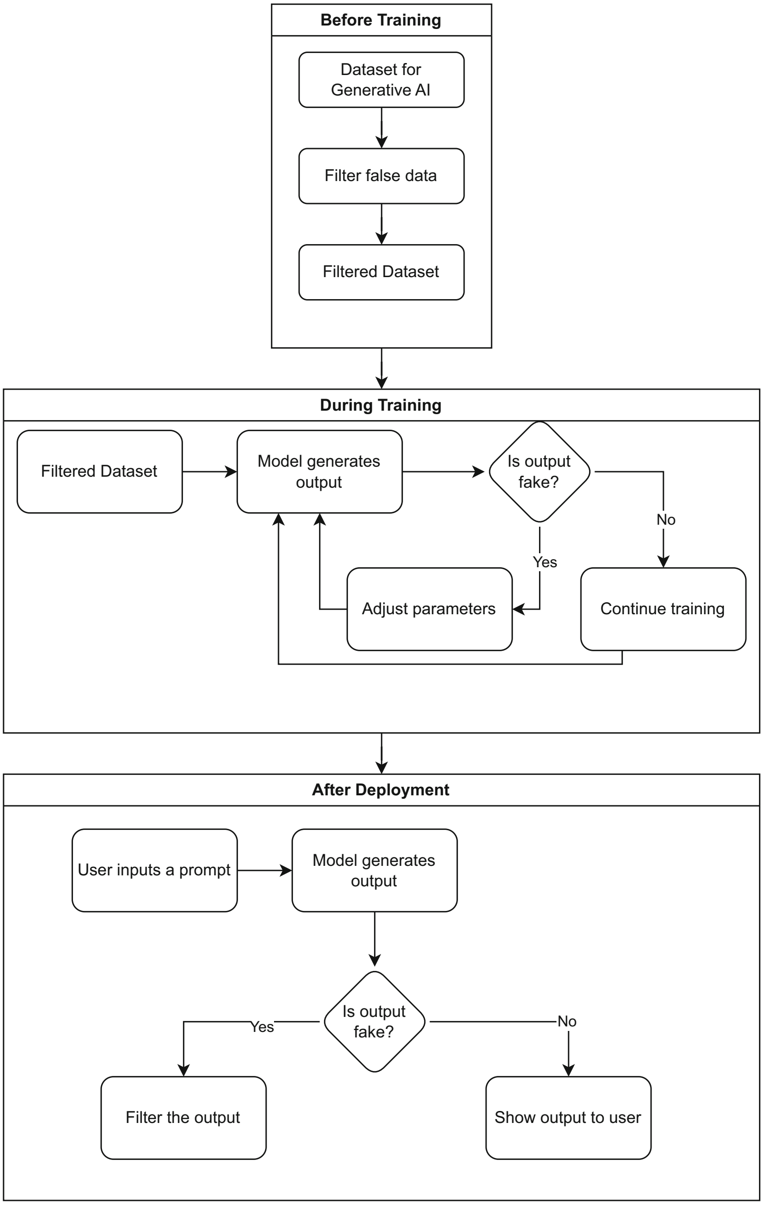

Fig. 11 Incorporation of Fake News Detector in Generative AI Models

- **Limited use of multi-modal data:** Fake news comes in all forms and sizes including as text and in images. Although a lot of research is done to detect its presence in textual data, still a lot of gap exists in detecting fake news in other media of communication.

8 Future directions

We had created this manuscript to enable further research in the areas so, this task could be added to the list of solved task of NLP. After creating this manuscript, there are several future research directions we want the reader to consider:

- **Improving cross-lingual support:** There is a need for fake news detection algorithms that can effectively handle multiple languages, especially in non-English speaking countries. Datasets are available as pointed out in Section 5, we just need to give attention to them. In this regard, we can explore to integrate Knowledge Bases of different languages in models that give satisfactory performance for other related languages, say Dutch knowledge base can be integrated in a model made primarily for English. We believe this area needs to be further explored.
- **Incorporation of multi-modal information:** As mentioned in Section 7 as well, multimodal nature of current era's data is a huge advantage we can make use of. Data comes through all forms of media be it written language, sound recordings, video recordings, sensory stimulation etc. Similarly, fake news propagates through all of these media and the researchers need to tackle that as well. Ensemble models can be trained to work on image attachments in a tweet along with the text of it. This would enable them to look at the samples from multiple dimensions and to get to the root of their meaning.
- **Adding more features:** We have observed from the studies that the features used are limited, often sticking to feature representations of textual data. Using more features can lead to better accuracy of models as their field of vision will be enhanced. These features could be comments on analysis, emotions of people, emoticons, and metadata of users. Often enough these features can conclusively identify a piece of information as fake e.g. presence of crossed fingers emoji etc. We believe instead of removing these features, we should inculcate them in a way that they help in meaningful classification. On the flip side, if the feature set becomes too large we need to look for efficient feature selection algorithms that can look for the best features to feed into the model.
- **Using sentiment analysis with fake news detection:** Although sentiment analysis of news has been done before, sentiment analysis of the user replies and the feedback of other users has not yet been done. We believe that such a study would help craft new features which can be used for the fake news detection task. Moreover, the sentiment analysis of user comments would give insight on how fake news affects the mass opinion - whether the mass follows the news or rejects it. We believe this area hasn't received the attention from the research community that it requires.
- **Transfer learning and active learning:** Researchers are exploring transfer learning and active learning techniques to make fake news detection algorithms more adaptable and efficient, especially in the context of fast-changing events like the COVID-19 pandemic. As the sources of news and types of news are evolving, we need models that use continuous learning processes to keep up with the changing landscape. In this regard adversarial networks need to be explored which keep on learning.

- **Development of explainable AI:** To ensure the transparency and accountability of fake news detection algorithms, there is a growing interest in developing explainable AI techniques [54, 62, 69] that can provide clear explanations and can be interpreted easily by observing the underlying models and decision-making processes. In a world of ever growing black box algorithms, we believe this would be a positive change.
- **Development of robust evaluation metrics:** In order to accurately measure the effectiveness of fake news detection algorithms, it is necessary to develop robust evaluation metrics that can account for various forms of bias and error.
- **Human-in-the-loop-approaches:** Human judgement can be integrated for fake news detection along with AI approaches. Humans can provide valuable context and insights that AI models miss.
- **Dataset diversity:** New datasets should be created that reflect evolving landscape of fake news. The datasets should cover various regions, and languages to improve accuracy of prediction models.
- **Psychological and sociological insights:** Collaboration should be done with psychologists to study psychological factors that make people more susceptible to fake news.
- **Study Spread through social networks:** Studies can be conducted on analyzing the spread of fake news on social media. Research should be conducted to find whether there is any difference in the way fake news spread as compared to real news.

9 Conclusion

In this survey paper we have reviewed popular techniques of NLP that are being used in the COVID-19 fake news detection. First the datasets used for fake news detection are described in detail. Then fake news detection task is reviewed on how it is moving forward and is suffering from model generalization problem. Datasets are explored, various studies are discovered and reviewed multiple approaches which deal with fake news detection using machine learning and deep learning techniques. Although, transformer based models are being widely used and provide state of the art results, hybrid ensembles surpass them. The review has unearthed the fact that people are generally unaware of the steps taken to minimize COVID-19 spread. Finally, we proposed a gap, in Section 8, which could be filled in future studies.

Funding The authors did not receive support from any organization for the submitted work.

Availability of data and material Data sharing not applicable to this article as no datasets were generated or analyzed during the current study

Declarations

Conflict of Interest Statement The authors declare that they have no conflict of interest.

References

1. Ghebreyesus TA (2020) Munich security conference 2. https://www.who.int/director-general/speeches/detail/munich-security-conference
2. Gambrell J, Karimi N (2020) In iran false belief a poison fights virus kills hundreds 2. https://www.pbs.org/newshour/world/in-iran-false-belief-a-poison-fights-virus-kills-hundreds

3. Satariano A, Alba D (2020) Burning cell towers, out of baseless fear they spread the virus 4. https://www.nytimes.com/2020/04/10/technology/coronavirus-5g-uk.html
4. Patwa P, Bhardwaj M, Guptha V, Kumari G, Sharma S, Pykl S, Das A, Ekbal A, Akhtar MS, Chakraborty T (2021) 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, Revised Selected Papers 1, Springer, pp 42–53. Accessed 8 Feb 2021
5. Borges TL (2022) Chrome extension for misinformation detection, Asian Journal For Convergence In Technology (AJCT) ISSN -2350-1146 8(3):6–11. https://doi.org/10.33130/AJCT.2022v08i03.002, https://www.asianssr.org/index.php/ajct/article/view/1240
6. Kenton JDM-WC, Toutanova LK (2019) Bert: Pre-training of deep bidirectional transformers for language understanding, In: Proceedings of NAACL-HLT, pp 4171–4186
7. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, Levy O, Lewis M, Zettlemoyer L, Stoyanov V, Roberta (2019) A robustly optimized bert pretraining approach, arXiv e-prints arXiv–1907
8. Yang Z, Dai Z, Yang Y, Carbonell J, Salakhutdinov RR, Le QV (2019) Xlnet: Generalized autoregressive pretraining for language understanding. Adv Neural Inform Process Syst 32
9. Pennington J, Socher R, Manning C (2014) GloVe: Global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), Association for Computational Linguistics, Doha, Qatar, pp 1532–1543. https://doi.org/10.3115/v1/D14-1162, https://aclanthology.org/D14-1162
10. Shu K, Sliva A, Wang S, Tang J, Liu H (2017) Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsl 19(1):22–36
11. Zhang X, Ghorbani AA (2020) An overview of online fake news: Characterization, detection, and discussion. Information Processing & Management 57(2):102025
12. Zhou X, Zafarani R (2020) A survey of fake news: Fundamental theories, detection methods, and opportunities. ACM Computing Surveys (CSUR) 53(5):1–40
13. Sharma K, Qian F, Jiang H, Ruchansky N, Zhang M, Liu Y (2019) Combating fake news: A survey on identification and mitigation techniques. ACM Trans Intell Syst Technol (TIST) 10(3):1–42
14. Bondielli A, Marcelloni F (2019) A survey on fake news and rumour detection techniques. Inf Sci 497:38–55
15. Detecting fake news with nlp (2023). https://medium.com/@Genyunus/detecting-fake-news-with-nlp-c893ec31dee8
16. Carrion-Alvarez D, Tijerina-Salina PX (2020) Fake news in covid-19: A perspective. Health promotion perspectives 10(4):290
17. Torales J, Barrios I, O'Higgins M, Almirón-Santacruz J, Gonzalez-Urbieta I, García O, Rios-González C, Castaldelli-Maia JM, Ventriglio A (2022) Covid-19 infodemic and depressive symptoms: The impact of the exposure to news about covid-19 on the general paraguayan population. J Affect Disord 298:599–603
18. Pereira Neto A, Ferreira EdC, Domingos RLAMT, Barbosa L, Vilharba BLdA, Dorneles FdS, Reis VSd, Souza ZAd, Graeff SV-B (2022) Assessment of the quality of information on covid-19 websites: an alternative for combating fake news, Saúde em Debate 46:30–46
19. Gisondi MA, Barber R, Faust JS, Raja A, Strehlow MC, Westafer LM, Gottlieb M (2022) A deadly infodemic: Social media and the power of covid-19 misinformation
20. Isaakidou M, Diomidous M (2022) The contribution of informatics to overcoming the covid-19 fake news outbreak by learning to navigate the infodemic. Stud Health Technol Inform 289:456–459
21. Wang X, Chao F, Yu G, Zhang K (2022) Factors influencing fake news rebuttal acceptance during the covid-19 pandemic and the moderating effect of cognitive ability. Comput Hum Behav 130:107174
22. Cahapay MB (2022) Covid-19 vaccine and vaccination misinformation and disinformation: Repositioning our role as educators in pandemic times. European J Environ Public Health 6(1) em0095
23. Williams NL, Wassler P, Ferdinand N (2022) Tourism and the covid-(mis) infodemic. J Travel Res 61(1):214–218
24. Whitehouse C, Weyde T, Madhyastha P, Komninos N (2022) Evaluation of fake news detection with knowledge-enhanced language models. In: Proceedings of the international AAAI conference on web and social media, vol 16, pp 1425–1429
25. Patwa P, Sharma S, Pykl S, Guptha V, Kumari G, Akhtar MS, Ekbal A, Das A, Chakraborty T (2021) Fighting an infodemic: Covid-19 fake news dataset. In: Combating online hostile posts in regional languages during emergency situation: first international workshop, Constraint 2021, Collocated with AAAI 2021, Virtual Event, Revised Selected Papers 1, Springer, pp 21–29. Accessed 8 Feb 2021
26. Cui L, Lee D (2020) Coaid: Covid-19 healthcare misinformation dataset, arXiv preprint arXiv:2006.00885
27. Paka WS, Bansal R, Kaushik A, Sengupta S, Chakraborty T (2021) Cross-sean: A cross-stitch semi-supervised neural attention model for covid-19 fake news detection. Appl Soft Comput 107:107393

28. Hayawi K, Shahriar S, Serhani MA, Taleb I, Mathew SS (2022) Anti-vax: a novel twitter dataset for covid-19 vaccine misinformation detection. Public Health 203:23–30
29. Raj C, Meel P (2022) Arcnn framework for multimodal infodemic detection. Neural Netw 146:36–68
30. Kim J, Aum J, Lee S, Jang Y, Park E, Choi D (2021) Fibvid: Comprehensive fake news diffusion dataset during the covid-19 period. Telematics Inform 64:101688
31. Zhou X, Mulay A, Ferrara E, Zafarani R (2020) Recovery: A multimodal repository for covid-19 news credibility research. In: Proceedings of the 29th ACM international conference on information & knowledge management, CIKM '20, Association for Computing Machinery, New York, NY, USA, pp 3205–3212. https://doi.org/10.1145/3340531.3412880
32. Wang WY (2017) "liar, liar pants on fire": A new benchmark dataset for fake news detection. In: Proceedings of the 55th annual meeting of the association for computational linguistics (volume 2: Short Papers), pp 422–426
33. Article scraping & curation - news (2023). https://newspaper.readthedocs.io
34. Yang Z, Yang D, Dyer C, He X, Smola A, Hovy E (2016) Hierarchical attention networks for document classification. In: Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies, pp 1480–1489
35. Shu K, Cui L, Wang S, Lee D, Liu H (2019) defend: Explainable fake news detection. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, pp 395–405
36. Poynter - poynter (2023). https://www.poynter.org
37. Khan JY (2019) A benchmark study on machine learning methods for fake news detection, 2. arXiv preprint arXiv:1905.04749
38. Newsguard - transparent reliability ratings for news and information sources (2023). https://www.newsguardtech.com/
39. Media bias/fact check - search and learn the bias of news media (2023). https://mediabiasfactcheck.com/
40. Malhotra R, Mahur A et al (2022) Covid-19 fake news detection system. In: 2022 12th International conference on cloud computing, data science & engineering (Confluence), IEEE, pp 428–433
41. Gundapu S, Mamidi R (2021) Transformer based automatic covid-19 fake news detection system, arXiv preprint arXiv:2101.00180
42. Thaipisutikul T, Shih TK, Enkhbat A, Aditya W, Shih H-C, Mongkolwat P (2022) Beyond fear go viral: A machine learning study on infodemic detection during covid-19 pandemic. In: 2022 14th International conference on knowledge and smart technology (KST), IEEE, pp 1–6
43. Mehta V, Mishra RK (2022) Machine learning based fake news detection on covid-19 tweets data. In: Proceedings of international conference on computational intelligence and data engineering: ICCIDE 2021, Springer, pp 89–96
44. Mazzeo V, Rapisarda A, Giuffrida G (2021) Detection of fake news on covid-19 on web search engines. Frontiers in physics 9:685730
45. Al-Ahmad B, Al-Zoubi A, Abu Khurma R, Aljarah I (2021) An evolutionary fake news detection method for covid-19 pandemic information, Symmetry 13 (6):1091
46. Covid-19 fake news dataset (2023). https://data.mendeley.com/datasets/zwfdmp5syg/1
47. Lovins JB (1968) Development of a stemming algorithm. Mech Trans Comput Linguist 11:22–31. https://api.semanticscholar.org/CorpusID:16628689
48. Mahlous AR, Al-Laith A (2021) Fake news detection in arabic tweets during the covid-19 pandemic. Int J Adv Comput Sci Appl 12(6):778–788
49. Wang Z, Zhao M, Chen Y, Song Y, Lan L (2021) A study of cantonese covid-19 fake news detection on social media. In: 2021 IEEE International conference on big data (Big Data), IEEE, pp 6052–6054
50. Shushkevich E, Alexandrov M, Cardiff J (2021) Covid-19 fake news detection: A survey. Computación y Sistemas 25(4):783–792
51. Wani A, Joshi I, Khandve S, Wagh V, Joshi R (2021) Evaluating deep learning approaches for covid19 fake news detection. In: Combating online hostile posts in regional languages during emergency situation: first international workshop, Constraint 2021, Collocated with AAAI 2021, Virtual Event, Revised Selected Papers 1, Springer, pp 153–163. Accessed 8 Feb 2021
52. Ameur MSH, Aliane H (2021) Aracovid19-mfh: Arabic covid-19 multi-label fake news & hate speech detection dataset. Procedia Computer Science 189:232–241
53. Heidari M, Zad S, Hajibabaee P, Malekzadeh M, HekmatiAthar S, Uzuner O, Jones JH, Bert model for fake news detection based on social bot activities in the covid-19 pandemic. In: (2021) IEEE 12th Annual ubiquitous computing, electronics & mobile communication conference (UEMCON). IEEE 2021:0103–0109
54. Kou Z, Shang L, Zhang Y, Wang D (2022) Hc-covid: A hierarchical crowdsource knowledge graph approach to explainable covid-19 misinformation detection, Proceedings of the ACM on Human-Computer Interaction 6 (GROUP) pp 1–25

55. Karnyoto AS, Sun C, Liu B, Wang X (2022) Augmentation and heterogeneous graph neural network for aaai2021-covid-19 fake news detection. Int J Mach Learn Cybern 13(7):2033–2043
56. Hande A, Puranik K, Priyadharshini R, Thavareesan S, Chakravarthi BR (2021) Evaluating pretrained transformer-based models for covid-19 fake news detection. In: 2021 5th International conference on computing methodologies and communication (ICCMC), IEEE, pp 766–772
57. Gautam A, Venktesh V, Masud S (2021) Fake news detection system using xlnet model with topic distributions: Constraint@ aaai2021 shared task. In: Combating online hostile posts in regional languages during emergency situation: first international workshop, Constraint 2021, Collocated with AAAI 2021, Virtual Event, Revised Selected Papers 1, Springer, pp 189–200. Accessed 8 Feb 2021
58. Karnyoto AS, Sun C, Liu B, Wang X (2022) Tb-bcg: Topic-based bart counterfeit generator for fake news detection. Mathematics 10(4):585
59. Karnyoto AS, Sun C, Liu B, Wang X (2022) Transfer learning and gru-crf augmentation for covid-19 fake news detection. Comput Sci Inf Syst 19(00):53–53
60. Gupta A, Sukumaran R, John K, Teki S (2021) Hostility detection and covid-19 fake news detection in social media, arXiv preprint arXiv:2101.05953
61. Bang Y, Ishii E, Cahyawijaya S, Ji Z, Fung P (2021) Model generalization on covid-19 fake news detection. In: Combating online hostile posts in regional languages during emergency situation: first international workshop, Constraint 2021, Collocated with AAAI 2021, Virtual Event, Revised Selected Papers 1, Springer, pp 128–140. Accessed 8 Feb 2021
62. Vijjali R, Potluri P, Kumar S, Teki S (2020) Two stage transformer model for covid-19 fake news detection and fact checking, arXiv preprint arXiv:2011.13253
63. Cheng M, Wang S, Yan X, Yang T, Wang W, Huang Z, Xiao X, Nazarian S, Bogdan P (2021) A covid-19 rumor dataset. Front Psychol 12:644801
64. Sarnovskỳ M, Maslej-Krešňáková V, Ivancová K (2022) Fake news detection related to the covid-19 in slovak language using deep learning methods. Acta Polytechnica Hungarica 19(2):43–57
65. Mookdarsanit P, Mookdarsanit L (2021) The covid-19 fake news detection in thai social texts. Bulletin of Electrical Engineering and Informatics 10(2):988–998
66. Mattern J, Qiao Y, Kerz E, Wiechmann D, Strohmaier M (2021) Fang-covid: A new large-scale benchmark dataset for fake news detection in german. In: Proceedings of the fourth workshop on fact extraction and VERification (FEVER), pp 78–91
67. Du J, Dou Y, Xia C, Cui L, Ma J, Philip SY (2021) Cross-lingual covid-19 fake news detection. In: 2021 International conference on data mining workshops (ICDMW), IEEE, 2021, pp 859–862
68. Cresci S, Di Pietro R, Petrocchi M, Spognardi A, Tesconi M (2017) The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race. In: Proceedings of the 26th international conference on world wide web companion, www ’17 companion, international world wide web conferences steering committee, Republic and Canton of Geneva, CHE, pp 963–972. https://doi.org/10.1145/3041021.3055135
69. De Magistris G, Russo S, Roma P, Starczewski JT, Napoli C (2022) An explainable fake news detector based on named entity recognition and stance classification applied to covid-19. Information 13(3):137
70. Biradar S, Saumya S, Chauhan A (2022) Combating the infodemic: Covid-19 induced fake news recognition in social media networks. Complex & Intell Syst pp 1–13
71. Glazkova A, Glazkov M, Trifonov T (2021) g2tmn at constraint@ aaai2021: exploiting ct-bert and ensembling learning for covid-19 fake news detection. In: Combating online hostile posts in regional languages during emergency situation: first international workshop, Constraint 2021, Collocated with AAAI 2021, Virtual Event, Revised Selected Papers 1, Springer, pp 116–127. Accessed 8 Feb 2021
72. Das SD, Basak A, Dutta S (2021) A heuristic-driven ensemble framework for covid-19 fake news detection. In: Combating online hostile posts in regional languages during emergency situation: first international workshop, Constraint 2021, Collocated with AAAI 2021, Virtual Event, Revised Selected Papers 1, Springer, pp 164–176. Accessed 8 Feb 2021
73. Malla S, Alphonse P (2022) Fake or real news about covid-19? pretrained transformer model to detect potential misleading news. The European Physical Journal Special Topics 231(18):3347–3356
74. Gonwirat S, Choompol A, Wichapa N (2022) A combined deep learning model based on the ideal distance weighting method for fake news detection. International Journal of Data and Network Science 6(2):347–354
75. Ahmed H, Traore I, Saad S (2018) Detecting opinion spams and fake news using text classification, Security and privacy 1(1):e9. https://onlinelibrary.wiley.com/doi/abs/10.1002/spy2.9, https://doi.org/10.1002/spy2.9
76. Zhang Y, Sun S, Galley M, Chen Y-C, Brockett C, Gao X, Gao J, Liu J, Dolan WB (2020) Dialogpt: Large-scale generative pre-training for conversational response generation. In: Proceedings of the 58th annual meeting of the association for computational linguistics: system demonstrations, pp 270–278

77. Faltl S, Schimpke M, Hackober C (2019) Ulmfit: state-of-the-art in text analysis, Internet: https://humboldtwi.github.io/blog/research/information_systems_1819/group4_ulmfit
78. Madani Y, Erritali M, Bouikhalene B (2021) Using artificial intelligence techniques for detecting covid-19 epidemic fake news in moroccan tweets. Results in Physics 25:104266
79. OpenAI, Gpt-4 technical report (2023). http://arxiv.org/abs/2303.08774 arXiv:2303.08774

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.