2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT) | 979-8-3503-7212-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/AIIoT58432.2024.10574600

Guardians of Truth: Exploring Techniques for Veracity Identification in social media during an Infodemic

P. Suthanthiradevi
Data Science and Business Systems, School of Computing,
SRM institute of Science and Technology, India
suthantp@srmist.edu.in

G Revathy
Department of Computer science and engineering
Vels Institute of science Technology and Advanced studies, India
grevathy19@gmail.com

Sathish Kumar P
Computer and Communication Engineering
Rajalakshmi institute of technology
India
sathishkumar.p@ritchennai.edu.in

Karthikeyan B
School of Computing,
SASTRA Deemed University,
India
mbalakarthi@gmail.com

Muthu Lakshmi.V
Department of Computer Science and Engineering,
St. Joseph's Institute of Technology,
India
hoditlabaffairs@stjosephs.ac.in

***Abstract*—With the recent advent of technology, social networks are accessible 24/7 using mobile devices. During covid-19 pandemic the propagation of misinformation are mostly related to the disease, its cures and prevention. With Twitter as the medium this information was spreaded to millions of people who lack domain specific knowledge. This can lead to bloster fear and direct damage to the people. This research focuses on identifying the veracity of the Twitter posts behind pandemic situation. The authors have proposed a Transformer Model using Bat Algorithm for identifying the fake news. This model is trained and evaluated using infodemic metadata. Our experimental results shows by which proposed model accurately detected the fake news and achieved 96.5% of f1-score.**

Keywords—infodemic, optimization, n-gram, BERT, Deep network, vectorization techniques

I. INTRODUCTION

Vocabulary into a searchable link, allowing anybody to click on it and see anything associated with it in real time. What role do hashtags play in providing a lifeline for people in the event of a global epidemic like COVID-19? COVID-19 is a well-known statistic that is now trending in the media including twitter. People from many walks of life are discussing varied thoughts and interests regarding their newfound time at home, such as: #Covid19, #Coronavirus, #StayHomeStaySafe and #indiafightscorona. "Fake along with effectively pertinent data promises in circulation which develop as a result of uncertainty, threat, and prospective damage, and which assist users generate knowledge also minimizes threat" is how rumors are distinguished from other types of news. The main impact of rumor in which misinformation is disguising healthy habits and encouraging erroneous practices, which contribute to widen the virus ultimately reaches poor physical and mental health outcomes among individuals. There have been several reports of catastrophes caused by these rumors all across the world [1]. Since the outbreak of COVID-19, there has been an explosion of misinformation and deception about the disease on social media and messaging platforms.

The COVID-19 epidemic to be an infodemic not a pandemic was officially reported by the World Health Organization [2]. Some of the information shared on social media is not accurate, it makes the users find trustworthy sources of media. On March 2020, the Australian researchers found that a tweet labeled as 'CORONAVIRUS' Chinese biological weapon' was common in tweet nearly 900 times. This tweet was retweeted 18,500 times and has 5 million views on the same day [3]. The huge quantity of false information on COVID-19 poses a risk towards the safety of the public. Ferrara et. al [4] Examined that social media might help us to understand how we're all dealing with this unprecedented global issue. Bots, or artificial identities, can, on the other hand, promote particular themes of conversation at risk of others on social media networks. Furthermore, in this paper, the author reviewed 43.3 million common languages tweeting about COVID-19 and observed initial signs of automated twitter accounts being used to promote conspiracy theories regarding politics in the United States, in stark contrast to people worried about public health. Evon et al. [5] refers to the fictional biological weapon as "Wuhan-400." It is also true that the coronavirus outbreak in 2020 will be centered in Wuhan, China. Jang et. al [6] proposed a model to learn about widespread viruses related tweets in NA (North America), particularly in Canada. The tweets relevant to that virus were evaluated via subject modeling and Aspect Based Sentiment Analysis (ABSA), moreover outcomes were discussed towards the health of society.

This paper aims to develop a model for detecting rumor information. Rumor spreading is usually detrimental to individuals and undoubtedly poses a threat to society as a whole. The research study demonstrates the importance of studying rumor information in social media and creating awareness to maintain a secure social life. By using deep learning algorithms with optimization [7], this research work provides a valuable analysis of social media text in order to prevent the diffusion of inaccurate information. The authors

979-8-3503-7212-0/24/$31.00 ©2024 IEEE

have proposed the Transformer Model using Binary Bat Algorithm (TMBBAT) is aimed to identify the rumor information without human intervention. Initially the authors performed EDA on benchmark COVID-19 metadata [12].
False claims about the transmission, prevention, and treatment of the sickness have led to people engaging in unclear and potentially hazardous activities such as applying bleach to their entire body and holding their breath for many minutes [14]. Section 2 discusses many investigations associated with the COVID-19 fake news detection. Section 3 illustrates the statistical analysis of the dataset. Section 4 explains the entire implementation of our proposed transformer model. Moreover, experimental outcomes of the proposed model are presented in Section 5. Brief conclusion of this research work is explained in section 6.

II. Literature Survey

Ramez Kouzy et. al [8] investigated 14 distinct popular hashtags and keywords connected to the COVID-19 outbreak to accomplish a Twitter search. Then, comparison on misunderstanding (fake news) of tweet posts among several researchers was verified and peer-reviewed resources, and analysed individual tweets for inaccuracy. To compare phrases and hashtags, as well as to identify individual tweets and account attributes, descriptive statistics were used. With the intention of acquiring untimely insights, Shahi et. al [9] undertook an exploratory investigation into the broadening, and substance of propaganda on Twitter approximately the issue of COVID-19. From the month of January to mid-July 2020, they gathered whole tweets that were mentioned in rulings on confirmed allegations relevant with COVID-19 by over 92 proficient confirmed organizations and shared such quantity with the community. This resulted in twitter 1500, with 1274 incorrect claims, 226 incompletely false claims as well. The average speed of tweet propagation is calculated by dividing the entire quantity of retweets for a tweet and the amount of days the tweets has been retweeted. The propagation speed of tweets can be calculated using the formula mentioned as equation 1.

$$Propagation\ speed\ of\ tweets\ (ps) = \frac{(number\ of\ tweets\ posted\ vs\ retweet\ count\ (rc))}{(Amount\ of\ days\ \ (nd))} \quad (1)$$

Where ps represents the speed of spreading, rc corresponds to the counting of retweets for each date and Nd signifies the number of days as whole.

Al-Zaman et. al [10] attempted to identify 5 essential elements of false news in social media relevant with COVID-19 among 125 Indian reports. Five important conclusions are drawn from the analysis on the following queries. Topics of false news are at first, health, religion, political, criminal, amusement, religious, and assorted. Singh et. al [11] determined the volume of interaction surrounding COVID-19 upon websites such as Twitter, which is besides the topics being discussed, the origin of this debate, common misconceptions regarding the disease along with the extent to which the discourse is connected to both significant or poor-quality materials from the web through used URLs connections. Furthermore, preliminary results show a significant spatially temporally relationship between the flow of data and newly discovered COVID-19 cases.

III. Exploratory Data Analysis

3.1. Dataset statistics

The proposed TMBBAT model has been implemented and analyzed with COVID-19 counterfeit report metadata. This data contains a corpus of real and fake information related to Coronavirus. Online social platforms like Twitter, Instagram, and Facebook are spreading misinformation during COVID-19 pandemic. Each data is labelled as real or fake based on their claims.
False claim: A clip of a healthcare professional combating COVID-19 inside an emergency room was created by NYT.
True claim: The video is real and Dr. Colleen Smith does work at the hospital, unlike the Tweet claims.

The following fig 1 shows the statistics of the fake news dataset. A total of 5600 data are identified as real whereas 5100 are fake news. Histogram shows that the number of real and fake data for training and testing are equally distributed so the dataset is balanced.

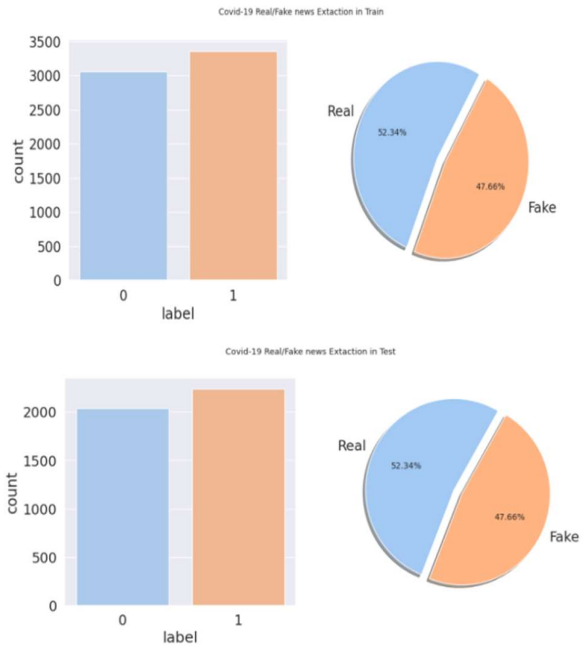

Fig. 1. Dataset visualization for training and testing data

3.2 N-Gram Analysis

The authors have analyzed the most distinct terms in the corpus with the aid of the Bag-Of-Words feature techniques. Figure illustrates the most common terms used in both real and fake news of COVID-19 data. In unigram token of 'test', 'confirm', 'report 'and 'health' are present in real news whereas 'claim', 'vaccine', 'trump' and 'death' are commonly used in fake news. The bigram tokens of the corpus. In bigram, the tokens of 'state report', 'active case'. çovid19 test' and 'mange isolation' are frequently used real bigrams 'donald trump', 'cure covid19', 'coronavirus vaccine' are fake bigrams. Trigram real tokens are 'total number test', 'number active case', 'daily update publish' and the fake trigrams of 'test positive covid 19','world health organization''and 'Facebook post claim'. Analyzing the n-gram has revealed that most of the tokens which are present in the real news are not similar in fake news. These unique

tokens are a very important source to identify the real and fake news.

IV. Methodology

Fig 2 depicts the proposed transformer model for detecting rumors in COVID-19 data. This section explains about data cleaning, text to feature vectorization, ML classifiers, also deep neural network models.

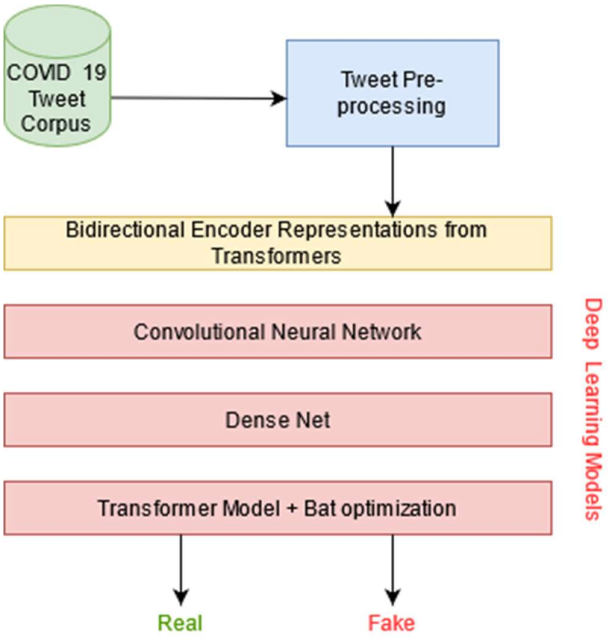

Fig. 2. Architecture of proposed TMBA Model

4.1 Preprocessing

COVID-19 dataset was collected from online social media, so it contains the special symbols such as hashtag, URLs and emoji etc. This uncleaned text data can affect the classifier performance. First, we have to remove URL, hashtags and @ symbols from the text in order to preserve the original meaning. The preprocessed text is converted to a feature vector with the help of the BERT embedding. The preprocessing rules are as follows

A) Word Tokenization (Tweet messages are converted to word tokens)

B) Removal of capitalization, hashtags, URL (#, URL link, @ symbols are removed)

C) Removal of stopwords (stopwords like 'we', 'have', 'and' is removed)

D) Apply Part of Speech (POS) tag (identify the correct verb and nouns.)

E) Convert POS tags to wordnet (Generates the lexical database)

F) Lemmatization (Identify variant form of same word)

4.2 Vectorization Techniques

Four frequently employed techniques for transforming texts into vector formats are shown in this section. The proposed research investigates the methods for exploring the vectorization of TF-IDF with stop words, TF-IDF without stop words, BERT word features and sentence features.

4.2.1 Bag of Words

A probabilistic linguistic framework called Bag-of-Words (BoW) has been employed to assess texts and content according to each word count. The sequence of words throughout the document is not taken into consideration by the design. A Python vocabulary comprising every keyword representing a single word and every parameter representing the quantity of occurrences that particular word occurs in a document might be used to create BoW.

4.2.2 TF-IDF with Stop words

The total amount of times a specific word appears within an individual file (term frequency) enhances the tf-idf score proportionately, whereas the total amount of texts within the database as a whole (inverse-document frequency) neutralizes the score. Every document is converted into rows via the tf-idf matrix, and each word is saved as a column vector. The tf-idf rating is determined by multiplying tf by idf described in eqn. (2).

$$TF - IDF(t,d,D) = TF\ (t,d) \times IDF(t,D) \quad (2)$$

Term-Frequency can be calculated using the formula mentioned as eqn. (3)

$$TF = \frac{Count_{t,d}}{Total\ count_d} \quad (3)$$

However, Inverse Term Frequency are evaluated using eqn. (4)

$$IDF(t,D) = \frac{i'}{df_t'} \quad (4)$$

4.2.3 TF-IDF without stop words

Here, the authors compute TF and IDF scores for each term in the documents, eliminating any stop keywords for assessment, while performing TF-IDF by utilizing stop phrases.

The following summarizes TF-IDF with no ending words implementation:

Tokenization: Give each document a unique set of words. To achieve this, the documents are divided into either bits or symbols.

- Determine the occurrence of every word within the file using term frequency (TF). To avoid biases favoring larger texts, TF might be computed utilizing a variety of techniques, including exponentially scaling rate, enhanced rate, and basic frequency of terms (the total amount of occurrences a word occurs in a text).
- Determine (IDF) for every phrase provides the logarithmic value for the entire number of docs multiplied by the total amount of docs that contain phrases commonly used to calculate IDF. This aids in compensating phrases which are more tolerant because they appear frequently in reports.
- Evaluate TF-IDF: To determine the TF-IDF score for each word in a specified document, perform multiplication of both TF and IDF.

4.2.4 BERT- Word features

Each input keyword in BERT (Bidirectional Encoder Representation from Transformers) uses a unique token, one of which is [HAVE]. When [HAVE] appears at the beginning as the input keyword order, the matching word it embedded is taken into consideration as being identical to the entire sentence. Put otherwise, the term "embedded" is employed as a word characteristic vector over phrase recognition. A document is treated as just one word when it is used as the input information for BERT. The document's vector of features is the word embedding of [HAVE], and it is possible to classify documents by building an algorithm and feeding it this characteristic vector.

4.2.5 BERT-Sentence features

It is intriguing to examine whether the prior processing of USER, HASHTAG, also URL tags may offer a benefit over zero preprocessing, several studies have shown how BERT Sentence embedded provides the optimal way to display text. The elimination of such phrases enhances feature extraction. This could be due to the fact that utilization of hashtags is highly common among social media subscribers, resulting in improvements in identifying trends and enhancing applied models' understanding of Coronavirus detection.

4.3 Deep Learning Models

4.3.1 Convolutional Neural Networks (CNN)

The authors have implemented the Convolution Neural Network [13] for detecting the fake news in the corpus. The N model consists of the BERT-embedding layer, two convolution layers, average pooling layers also FC layer. With BERT sentence embedding, the input sentences are transformed to bert-input (in_id, in_mask and in_segment). The matrix of BERT-sentence is fed to 1D convolutional with dimensions 3,5 and filter size is 128. The dimensionality of the matrix is reduced by max_pooling layer. The reduced matrix output is fed to the FC layer which is connected with the SoftMax classifier.

4.3.2 Deep Network

The deep network model is trained on the BERT sentence embeddings [16]. It is a pre-trained embedding method that maps every token in the covid-19 fake news data to the similarly labeled data. The input tokens are converted to the embedding vector. The output of the BERT layer is based on the sentence similarity vector passed to the input dense layer. Final output layer with 'sigmoid' classifier has identified the real/fake labels.

4.3.3 Proposed Transformer Model using Binary Bat Algorithm

Transformer model is used to process the sequential data in parallel. All input sentences are passed parallelly. The transformer components are

(a) Input Embedding - To map every token of the data to similar meaning tokens.

(b) Positional Encoding- It generates the word vector with positional information i.e. context.

(c) Encoding and Decoding Layer- It focuses on the important tokens in the corpus. Attention vector captures the contextual relationship between the tokens and sentences. Transforms the attention vector into the digestive format.

(d) Linear Layer- Expand the vector size into count of tokens in the data.

(e) SoftMax-Transforms the vector into a probability distribution.

To pre-train the bidirectional representation of the unlabeled data, BERT Natural Language Processing model has been developed. This model comprises two basic concepts (i) Atten mechanism can process the long sentences effectively. (ii) Bidirectional training has a deeper sense of text context and predicts the meaning of the token very effectively. Every parameter of BERT is finely tuned with labeled data for detecting fake news. This base BERT model comprises 12 encoder layers, one drop out layer, two linear networks and SoftMax activation function. The longest sentence possible is 512. SoftMax classifier model is to classify the probability of the real or fake labels.

The transformer uses an attention mechanism to identify the context of the sentence [15]. The BERT model is fine-tuned by setting the hyper parameter values in the proposed model. This model is based on the binary bat algorithm variants named as TM-BBA in Table 1. The BBA optimizer is used to find the optimal parameter settings for tuning the BERT model. The authors have considered the parameters such as number of epochs, batch size, learning rate and learning rate warmup steps. The proposed deep neural network of TM-BBA model achieves better accuracy of 95.64% for detecting fake news.

Table 1. Optimal Parameter solution for TM-BBA Model

Parameter	Deep Network	TM-BBA
Number of epochs	10	5
Batch size	16	8
Learning rate	le-4	2e-6
Learning rate warmup steps	10000	8000

V. EXPERIMENTAL RESULTS

The experimental results of proposed TM-BBA model aids to classify the COVID-19 tweets as real and fake using the transformer model. The deep learning model with fine tune parameter optimization achieves better accuracy as compared to other deep learning models. This model is trained on a GPU based system. The python deep learning packages such as transformers, pytorch, TensorFlow hub, keras, pandas and NumPy are used to implement the proposed model.

5.1 Pre-processing

The data corpus contains the collection of COVID-19 fake news information. This data contains noisy information which leads to affecting the model performance. The authors have applied preprocessed rules to clean the data corpus. The results of the preprocessed text are converted to vectors with the aid of the BERT.

5.2 Sentence BERT

The preprocessed tweets are converted to vectors using sentence BERT (Fig.3). It generates the vector for every token in the data how much the token is related to every token in the sentence. Sentence Bert uses the triplet and siamese network structure to identify the semantic meaning of the sentence. The output of the BERT sentence features are fed into the various transformer models to identify the real and fake news.

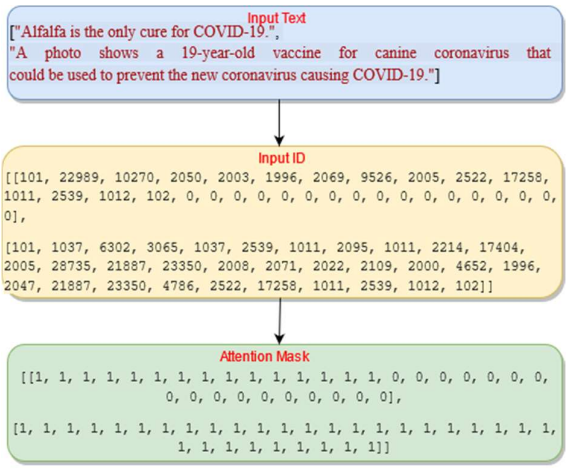

Fig. 3. Sentence Bert Output

5.3 Performance of Vectorization Techniques

The primary objective of this research is to establish benchmarks along with assess the efficacy of fake news detection and classification with reference to coronavirus. The authors utilized COVID Senti datasets and applied machine learning, neural network technique, Dense Net, transformer, and Bat optimization model. It then evaluates the results using performance assessment metrics such as F1-score, recall, accuracy, and precision. The model's ability to produce the most effective classification outcome is attested to by the meticulously chosen conventional performance measures. During the testing process, different vectorization strategies such as Bag of Words, TF-IDF with and without stop words, BERT word and sentence features are compared to the experimental results of SVM, NB, and Decision tree in Table 2.

Table 2. Vectorization Techniques vs Model

Vectorization Techniques	SVM	NB	Decision Tree
Bag of Words	91.85	89.07	88.46
TF-IDF with Stop words	94.11	90.82	87.76
TF-IDF without stop words	93.97	90.89	87.69
Bert- Word features	92.80	79.95	80.00
Bert-Sentence features	91.36	80.00	79.37

5.4 Performance of the proposed TM-BBA model

The aforementioned research primary goal is to locate fraudulent information on social networks. The proposed model is trained COVID-19 Fake news Detection dataset. The pre-trained BERT models are implemented with the aid of the pre-trained values. The hyper parameter values are fine-tuned by using the optimization algorithm. The TMBA model's performance is evaluated with the aid of the evaluation metrics and shown in table 3. The results prove that a transformer model with fine tune optimization parameter values achieves better performance. The weighted average of the F1-score for BERT sentence embedding features with dense networks achieves nearly 0.9 % better score than the convolutional neural network. The TMBA model achieves a 0.4% increase in f1-score as compared to the dense network.

Table 3. Evaluation metrics of various transformer model

Transformer Models	Accuracy	Precision	Recall	F1-score
CNN	0.83	0.85	0.83	0.83
Dense layer	0.92	0.93	0.92	0.92
Proposed TMBBA	0.96	0.95	0.96	0.96

The proposed model is compared to the existing model. The author [17] evaluated the model using CNN and LSTM deep learning models. To enhance the performance of the proposed model we combined the CNN with optimization methods. In Table 4, the final results prove that the proposed transformer model with the optimization parameter values detect the fake news text better as compared to the other deep learning models.

TABLE 4. COMPARISON OF PREVIOUS WORK

Transformer Models	Accuracy	Precision	Recall	F1-score
CNN [17]	92.99	92.95	92.90	92.90
LSTM	92.90	92.92	92.90	92.89
Proposed TMBBA	96.12	95.83	96.10	96.5

CONCLUSION

The infodemic dataset is analyzed using deep neural networks with transformer models. Initially we explored the statistics of the real and fake data using N-gram analysis. The proposed TMBA model is to predict the fake news content in the corpus. COVID-19 fake news data are experimented with

different deep learning models using pre-trained embedding parameter values. The TMBA model is fine-tuned with the BERT embedding layer hyper parameters based on the Bat optimization algorithm. The results prove that the TMBA model is identifying the fake news with an improved F1-Score of 96.5%. In future the proposed model is evaluated using different evaluation metrics.

REFERENCES

[1] Lisa Singh, Shweta Bansal, Leticia Bode, Ceren Budak, Guangqing Chi, KornraphopKawintiranon, Colton Padden, Rebecca Vanarsdall, Emily Vraga, Yanchen Wang "A _rst look at COVID-19 information and misinformation sharing on Twitter", arXiv:2003.13907v1 [cs.SI] 31 Mar 2020.

[2] Ferrara, E., 2020. # covid-19 on Twitter: Bots, conspiracies, and social media activism. arXiv preprint arXiv:2004.09531.

[3] Dhiman, P., Kaur, A., Iwendi, C., & Mohan, S. K. (2023). A scientometric analysis of deep learning approaches for detecting fake news. Electronics, 12(4), 948.

[4] https://www.newsbytesapp.com/news/science/twitter-will-now-flag-covid-19-rumors-with-labels/story,Shubham Sharma 12 May 2020

[5] Analysis of millions of coronavirus tweets shows 'the whole world is sad' Ben Guarino March 17, 2020

[6] https://www.politico.com/news/2020/06/02/trump-supporters-on-twitter-spread-covid-19-rumors-about-china-296808,Mark Scoot,06/02/2020 04:19 PM EDT

[7] Chen, M. Y., Lai, Y. W., & Lian, J. W. (2023). Using deep learning models to detect fake news about COVID-19. ACM Transactions on Internet Technology, 23(2), 1-23.

[8] Evon, D., 2020. Was Coronavirus Predicted in a 1981 Dean Koontz Novel? URL: https://www. snopes. com/fact-check/dean koontz-predicted-coronavirus.

[9] Jang, H., Rempel, E., Carenini, G. and Janjua, N., 2020. Exploratory analysis of COVID-19-related tweets in north America to inform public health institutes. arXiv preprint arXiv:2007.02452.

[10] Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C.M., Brugnoli, E., Schmidt, A.L., Zola, P., Zollo, F. and Scala, A., 2020. The covid-19 social media infodemic. arXiv preprint arXiv:2003.05004.

[11] Kouzy R, Abi Jaoude J, Kraitem A, El Alam MB, Karam B, Adib E, Zarka J, Traboulsi C, Akl EW, Baddour K. Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter. Cureus. 2020 Mar 13;12(3): e7255. doi: 10.7759/cureus.7255. PMID: 32292669; PMCID: PMC7152572.

[12] Shahi, G. K., Dirkson, A., & Majchrzak, T. A. (2021). An exploratory study of COVID-19 misinformation on Twitter. Online social networks and media, 22, 100104.https://doi.org/10.1016/j.osnem.2020.100104

[13] Al-Zaman, M.S. 2021. COVID-19-Related Social Media Fake News in India. Journalism and Media 2: 100–114. https://doi.org/10.3390/ journalmedia2010007

[14] Qu, Z., Meng, Y., Muhammad, G., & Tiwari, P. (2024). QMFND: A quantum multimodal fusion-based fake news detection model for social media. Information Fusion, 104, 102172.

[15] Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. (2019). NAACL-HLT. https://constraint-shared-task-2021.github.io/

[16] LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. (2015). Nature, 521(7553), pp. 436-444

[17] R. Malhotra, A. Mahur and Achint, "COVID-19 Fake News Detection System," 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2022, pp. 428-433.