Transfer Learning Model for Disrupting Misinformation During a COVID-19 Pandemic

Rini Raju

Shova Bhandari

Sofia A.Mohamud

Ebrima N Ceesay

(sbhanda7, smohamu2, rriniraj, eceesay2)@gmu.edu
George Mason University

2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC) | 978-1-6654-1490-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/CCWC51732.2021.9376066

***Abstract*—In 2020, the COVID-19 pandemic changed the world significantly, and it is critical to have reliable online information about this virus. However, disinformation can have a negative effect on public opinion and can put the lives of millions in danger by ignoring the crucial precautions. People worldwide post their ideas about the coronavirus every second and create a rich source of information. In this work, we introduce an advanced natural language processing model to classify public opinion about the virus, which can help health organizations to take immediate actions to stop the spread of the virus by removing misinformation from online platforms. We introduce a new model with high classification accuracy to extract deep contextual information from online coronavirus comments based on main COVID-19 topics and use a robust model for sentiment classification based on more than twenty different datasets to detect the tweet's text, which contains misinformation. The new model can generate reports about tweets that contain misinformation to the states requiring emergency precautions to stop the virus's spread by removing the detected comments from their online platforms. Also, The new model can detect misinformation and prevent fake news by increasing public awareness about COVID-19.**

***Index Terms*—Machine Learning, Misinformation, COVID-19, Natural Language processing**

I. Introduction

Since the flu season is approaching, stress on the healthcare system, especially hospitals, will be significantly impacted by the overlap of COVID-19 and flu infections. Also, the opening of universities might increase the spread of COVID-19 as students from different parts of the country will come together [1]. Figure 1, provided by the CDC, shows several coronavirus cases in April when social distancing was not taken seriously by public opinion. However, moving to May, there was a 50% reduction in daily contacts, and in June, there was about 75% reduction in daily contacts. So the public opinion about the pandemic is critical and can affect the death toll significantly.

Several conspiracy theories and misinformation regarding public health concerns have emerged over the years. There has been a surge of misinformation and conspiracy theory in response to the COVID-19 pandemic [2]. Believers of conspiracy theories are less keen to follow any health advice [3]. Beliefs in a COVID-19 conspiracy theory make it hard for the state and federal government to control the COVID-19 [4].

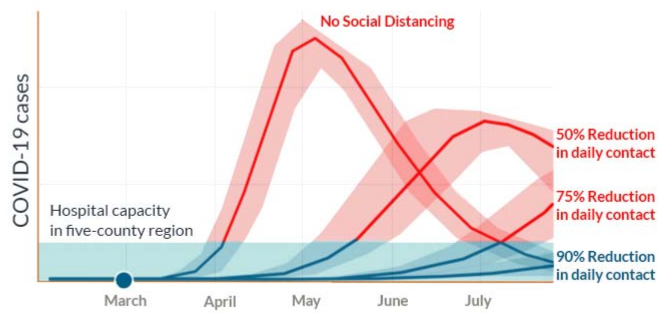

Fig. 1. The effect of Social distancing acceptance in public opinion

Romer & Jamieson survey of 840 individuals shows that :

- More than 37% believe that the COVID-19 virus is human-made created by a Chinese in a lab [4].
- More than 32 % believe that the CDC inflated the danger of COVID-19 and death to affect political events [4].

According to Laato et al. [5] research model, "a person's trust in online information and perceived information overload are strong predictors of unverified information sharing." According to Pennycook et al. [6] investigations of 1600 individuals, people share fake news without knowing the truthfulness and value of the news or information.

The primary threat regarding misinformation is that social media users tend to share fake COVID news more than facts-based news leading to conflicting and poor decision making [7]. According to the survey of 483 participants, conspiracy, political, and religious misinformation beliefs about COVID-19 impairs decision making and have a negative impact on individual responses [8] [9]. Nevertheless, social media platforms like Facebook and Twitter have taken significant actions "to try to limit the proliferation of disastrous misinformation regarding COVID-19 by removing fact-checked false and potentially harmful information" [8] [10]. Social bots can spread misinformation about Coronavirus, and expert bot detection

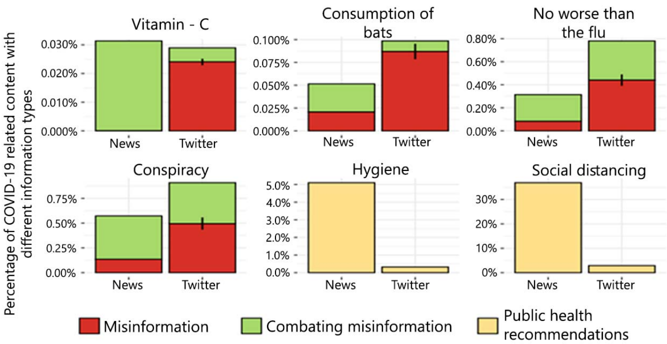

Fig. 2. COVID-19 misinformation is distributed more on Twitter as compared to traditional media [17]

methods [11] [12] [13]can be very helpful to detect these types of bots based on the deep context of online comments in the global pandemic. Summarizing the COVID-19 comments by machine learning models can show people's views about global pandemic [14] [15] and can detect misinformation.

This work creates a new model for binary classification of COVID-19 comments, which detects tweets containing misinformation. Each component in the new model is chosen based on the best classification accuracy. Also, memory usage and computational time optimization in the training phase are considered in model evaluation. All the COVID-19 comments on social media platforms are unlabeled texts, and it is costly and time consuming to label all comments. So We use new transfer learning approach based on studies by [16] for semantic extraction of online comments.

II. Data set

In this research, we use COVID-19 Open Research Datasets that were collected by the NIH, CDC, White House, and George Town university [18] as the main source for factual information. Also, it includes 59,000 scholarly articles that provide information on the coronavirus [18]. Kaggle data sets in this research have USA COVID-19 data set as well [18]. "Uncover COVID-19 challenge" is another data set by Roche Data Science Coalition [18]. We also use the COVID-19 Twitter chatter dataset [19] and choose English tweets for the sentiment classification in this research.

Also, there are various Twitter data sets about social interactions of Twitter users and their behaviors about the COVID-19 pandemic [20], which is used in this research. The main reason that we use different types of COVID-19 datasets is that having a fair amount of COVID-19 tweets in each category of topics, which is used in this research. Figure 2 displays disinformation on social media compared with traditional news media. As can be seen, social media platforms have more percentage of misinformation sharing compared with the news COVID-19 [21].

III. Methods

According to work by Heidari et al. [11], this research creates a new model that can classify unlabeled data of Covid-19 tweets' texts with high accuracy for misinformation based

TABLE I
TEXTBLOB RESULTS FOR SENTIMENT CLASSIFICATION OF COVID-19

Topics	tweet sentiments		
	positive	negative	neutral
flu	38%	20%	42%
face-mask	42%	10%	48%
politics	4%	8%	88%

on the BERT model. Before the sentiment classification of public opinion, we apply LDA topic modeling to collect all related tweets to covid-19 topics. LDA topic modeling identifies specific collocated keywords used when commenting on COVID-19 on social media platforms. It will help us identify the number of people who believe in the controversy related to COVID, such as:

- COVID-19 is related to politics, and it is human-made (politics)
- Face mask is not effective in Virus prevention(face mask)
- It is similar to the flu(Flu)

Three main categories of COVID-19 related topics are Flu, Face mask, and Politics, which are selected by LDA topic modeling. So we collect an equal number of tweets related to each topic, 500,000 tweets in each category from different datasets. We want to introduce a model that can clearly show the sentiment and the context of public opinion about the correlation of the COVID-19 virus and these three Topics. The CDC, governmental agencies, and health organization can use the results of this research to adopt health policies to stop the spread of the virus by increasing public awareness and detecting the tweets that contain misinformation about the virus. Before we introduce a new model, We select the sentiment classification model for the COVID-19 tweets. Then we create a labeled dataset by using human annotators as ground truth for this research and evaluate our new model to detect tweets that contain misinformation.

A. TextBlob for sentiment analysis of public opinion based on COVID-19 Tweets

After Collecting the tweets based on the LDA topic modeling, at first, for sentiment analysis of online comments, we examine TextBlob performance. We used the TextBlob to label the Covid-19 tweets in three categories: Positive, negative, and neutral comments. The TextBlob is based on the Naive Bayes classifier, a supervised learning method, and it assigns 1 for positive tweets, -1 for negative tweets, and 0 for neutral tweets. The class label is assigned to the tweet's text based on the sentiment label with the most significant sentiment value.

$$T_i = \begin{cases} \text{Positive if } TS_i > 0 \\ \text{Negative if } TS_i < 0 \\ \text{Neutral if } TS_i = 0 \end{cases} \tag{1}$$

Table I shows the results of TextBlob based on the tweets filtered by topic modeling. Tweets are categorized into three different categories "flu," "Face Masks," and "politics." In each category, all related tweets are collected. Since the tweet's

TABLE II
TF-IDF RESULTS FOR SENTIMENT CLASSIFICATION

Topics	tweet sentiments		
	positive	negative	neutral
flu	38%	22%	40%
face-mask	10%	30%	60%
politics	12%	6%	82%

TABLE III
BERT RESULTS FOR SENTIMENT CLASSIFICATION OF COVID-19

Topics	tweet sentiments		
	positive	negative	neutral
flu	70%	20%	10%
face-mask	35%	45%	20%
politics	35%	40%	25%

text about covid-19 are unlabelled data, as can be seen, the TextBlob is not a robust classification model since many tweets are categorized as neutral. The number of positive tweets about the (Covid-19, flu) and (COVID-19, Face mask) is more than the number of positive tweets about the(COVID-19, Politics). Also, nearly all the comments about politics and COVID-19 are categorized as neutral. Since many tweets are classified as neutral comments, it is hard to understand the public opinion about different topics. So we need a more advanced model for sentiment classification for coronavirus tweets.

B. TF-IDF for sentiment classification of COVID-19 Tweets

Based on the previous work [22] we built a customized TF-IDF model for tweet sentiment classification. A new customized TF-IDF method is built for the words in each tweet's text which are more relevant to the topics extracted by LDA topic modeling. A word is considered an important word if it appears several times in one document(TF), and it appears less frequently in all relevant documents. We create a corpus-based on information extracted from the 48 coronavirus articles, which is available online in the Kaggle dataset. Then We calculate the multiplication of term frequency in Inverse-Document-Frequency. The following formula is used to find similarity for each tweet's text to positive and negative COVID-19 comments based on LDA topic modeling.

$$sim(d_j, q) = \frac{\sum_{i=1} w_{i,j} \times w_{i,q}}{\sqrt{\sum_{i=1} w_{i,j}^2} \times \sqrt{\sum_{j=1} w_{i,q}^2}}$$

Table II shows the sentiment classification results from the TF-IDF method. The number of neutral tweets for each topic is very high compared to the TextBlob model, and it reduces the chance of observing the people's opinion about COVID-19 with three selected topics. Also, the TF-IDF model's high computation time for sentiment classification of COVID-19 tweets affects the algorithm performance. So We need a more advanced natural language processing model to extract public opinion based on the deep context of COVID-19 tweets and increase the sentiment classification accuracy by reducing the number of neutral tweets. In the next section, we use the transfer learning approach for sentiment classification.

C. BERT for sentiment classification of COVID-19 Tweets

In this research, we use Bidirectional Encoder Representations from Transformers (BERT) [23] [24] for sentiment classification based on the deep context of the tweet's text. The successful sentiment classification of the BERT for several applications has been studied before, and in this work, we use the transformer model for sentiment classification of the COVID-19 related tweets. The transformer BERT's bidirectional training is a strong attention model to classify COVID-19 tweet's text into negative, positive, and neutral posts. The BERT model's main advantages are parallel computation ability and training unlabeled tweets of COVID-19 in social media platforms.

For sentiment classification on the tweet's text, one approach can be using a feature-based solution such as ELMO [25], which is a deep contextual word representation. Another solution is a fine-tuning approach, such as Generative pre-trained transformer(GPT) [26], but they are not strong models for sentiment classification of tweets compared with using BERT [23]. ELMO and GPT both useful in creating language representation, but BERT is a pre-training of a deep bidirectional transformer, a powerful model for language understanding. We use BERT for single sentence classification task based on SST-2 [27]. SST-2 is Stanford Sentiment Treebank of binary classification of single sentences based on movie reviews.

Table III shows the sentiment classification results by BERT model. As can be seen, the number of tweets classified as neutral is reduced significantly, and BERT can classify positive and negative tweets compared with two previous models (TextBlob, TF-IDF) with more accuracy. The BERT model can extract the context from COVID-19 tweets with more precision since it considers the backward and forward history of each word in each tweet's text. Reducing the number of neutral tweets helps us to have deep context extraction about public opinion about the virus on online platforms. Although the BERT model provides higher accuracy for sentiment classification in comparison with previous methods, Since the training of the BERT is computationally expensive, in the next section, we use models for the analysis of public opinion, which can improve the training computation time and provide the accuracy in sentiment classification as well. Figure3 shows a comparison of sentiment classification of COVID-19 tweets based on TF-IDF, TextBlob and BERT. As can be seen, the BERT sentiment classification accuracy outperformed previous models in three different topics by providing the minimum number of neutral comments.

D. RoBERTa for sentiment classification of COVID-19 Tweets

This section examines two other advanced natural language processing models' performance to extract the public opinion about COVID-19 comments on three topics(Flu, Face mask, Politics). BERT model has several hyperparameters that can be optimized to improve algorithm performance significantly.

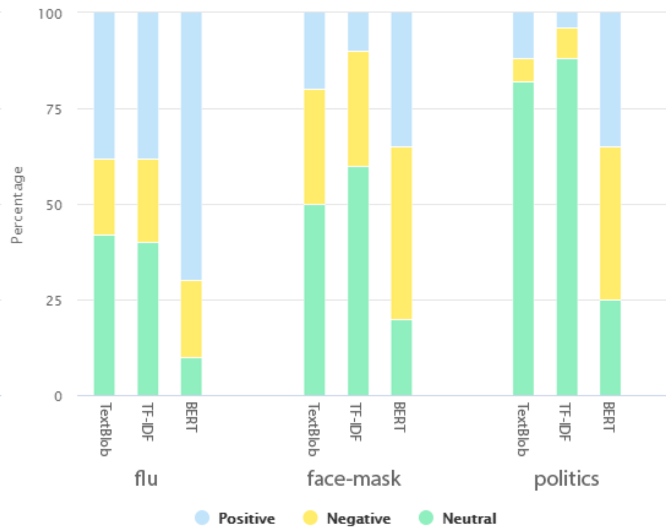

Fig. 3. Comparison of Sentiment Classification of COVID-19 tweets based on TF-IDF, TextBlob and BERT

TABLE IV
F1-SCORE FOR BERT BASE MODEL BASED ON DIFFERENT CONFIGURATION OF THE ALGORITHM [28]

	dynamic masking	static masking	without NSP	with NSP
f1-score for BERT base model	92.9	92.5	92.9	92.8

In this section, we use RoBERTa [28](Robustly Optimized BERT Approach) model; we apply changes based on several hyperparameters that can reduce the computation time for sentiment classification of COVID-19 tweets. In the BERT base model for masking each tweet's text, static masking is used in every training phase. However, using Dynamic masking based on the tweet's text in each training phase is more efficient. Table IV shows how dynamic masking of the tweet's text sequence in the RoBERTa model can improve the performance of the BERT base model. So in this work, we use dynamic masking for all COVID-19 tweets.

TABLE V
ROBERTA PERFORMANCE BASED ON DIFFERENT BATCH SIZES

	256	2K	8K
RoBERTa f1 score for different batch size	92.7	92.9	92.8

TABLE VI
ROBERTA RESULTS FOR SENTIMENT CLASSIFICATION OF COVID-19 TWEETS [28]

Topics	tweet sentiments		
	positive	negative	neutral
flu	75%	20%	5%
face-mask	35%	55%	10%
politics	35%	45%	20%

TABLE VII
ALBERT VS BERT BASE MODEL [33]

model	Parameters	Layers	Hidden	Embedding	Parameter sharing
BERT-base	108M	12	768	768	False
ALBERT-base	12M	12	768	128	False

Also, in the BERT base model, the next sentence prediction tasks(NSP) loss have a negative effect on downstream tasks in natural language processing; the RoBERTa model improves the BERT model based on the results in Table IV by removing NSP loss and shows that how the algorithm performance can be affected by NSP loss. In this work, We remove the next sentence prediction loss to improve our sentiment classification model. The batch size in the BERT model's training plays an essential role in algorithm accuracy and can affect the sentiment classification of the COVID-19 tweets; we changed the batch size based on the Optimized RoBERTa batch size for the BERT base model. Table V shows that how RoBERTa can improve the NLP tasks by changing the batch size in the time of training. Also, larger batch sizes help to use parallel computation and distributed systems more efficiently in sentiment classification. Since BERT is trained based on Book corpus [29] and English Wikipedia, in this work based on the RoBERTa model, we use a more comprehensive source of language corpus for our sentiment classification model, namely: CC-News [30] which is based on the news data sets, OPEN WEBTEXT [31] which is extracted web contents, STORIES datasets [32]. We also used the original datasets of the BERT base model BOOK corpus and English Wikipedia. The data used to train RoBERTa in this work is 160 GB, which covers different domains. The results of sentiment classification of COVID-19 tweets based on the RoBERTa model are shown in Table VI. As can be seen, the percentage of neutral tweets reduced compared to when we use the BERT model. The computation time for training the COVID-19 tweets for sentiment classification reduced significantly compared with the BERT base model.

E. ALBERT for sentiment classification of COVID-19 tweets

Although we improve the sentiment classification of COVID-19 tweets by RoBERTa, the training model and the memory consumption can be optimized by doing new parameter reduction techniques in the BERT base model. In this section, we examine the ALBERT [33] model for sentiment classification of COVID-19 tweets. ALBERT model uses two main parameter reduction techniques to optimize the BERT base model. First, the ALBERT model divides the larger embedding of the vocabulary into two small matrices. This division provides more flexibility for the hidden layers' size and makes it independent of vocabulary embedding size. Second, cross-layer parameter sharing techniques prevent increasing the number of parameters when the network's depth increases in the training phase. These two BERT base model changes are crucial for our sentiment classification of the COVID-19

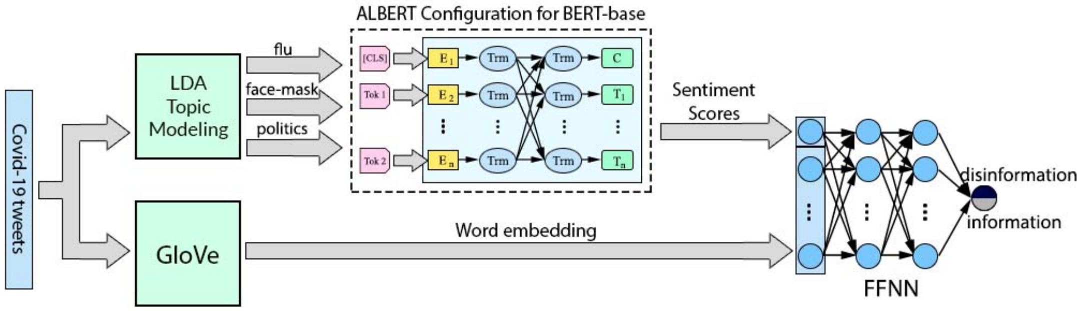

Fig. 4. The new model for disinformation detection

TABLE VIII
ALBERT RESULTS FOR SENTIMENT CLASSIFICATION OF COVID-19 TWEETS

Topics	tweet sentiments		
	positive	negative	neutral
flu	75%	21%	4%
face-mask	35%	60%	5%
politics	37%	45%	18%

tweets and our new model for misinformation detection in this research. Table VII shows different parameters that we consider to improve sentiment classification by the ALBERT model. Table VIII shows the sentiment classification results by using ALBERT model. As can be seen, the number of neutral tweets is less than the RoBERTa model. The main improvement here is the time of the training model for sentiment classification of the COVID-19 tweets, which is significantly reduced by using the ALBERT model instead of RoBERTa model. Since during the pandemic, people post many online comments about COVID-19, the time of the training model for sentiment classification is very important because of the huge number of generated tweets. In this work, we choose the best model for sentiment classification based on accuracy, computation time, model training time, memory usage, and parallel computation ability.

We use ensemble learning in the studies by Heidari et al. [34] to design a final model for misinformation detection. Figure 4 shows our new model for tweet misinformation detection. In the first phase of the model, all tweets are categorized by LDA topic modeling. Then we assign different sentiment scores to each tweet based on the ALBERT model. Since there is a possibility that one tweet can be categorized in more than one topic, three sentiment scores for each topic will be assigned to each tweet. Based on sentiment classification accuracy, we choose ALBERT as a sentiment classification component in this model. We use 120 human annotators who are provided by Amazon Sagemaker [35] to make an annotation for 600,000 COVID-19 tweets. Each annotator needs to label a tweet's text as information or misinformation based on CDC reports. This dataset is used as ground truth to evaluate the accuracy of our new model. As can be seen in figure 4 ALBERT model generates sentiment scores for COVID-19 tweets, which is passed as an input of the Feedforward neural network. In addition to sentiment scores for each tweet's text, we consider the tweet's text as input to the final neural network classifier. We use GloVe(Global Vectors) [36], which is statistics of the global English corpus for each token in a tweet's text in the word embedding phase. Sentiment scores of each tweet's text and word embedding of the tweet's text feed the final classifier, a feed-forward neural network model. The final neural network model has two hidden layers. The model classifies the tweets into two categories of information or misinformation. Table IX shows the performance of the neural network model in the detection of misinformation as a final classifier in our new model based on different sentiment classification techniques to generate sentiment scores. Choosing ALBERT to generate sentiment scores for COVID-19 tweets makes our new model achieve high accuracy of 0.91 by the final classifier to detect the misinformation. Also, RoBERTa can be chosen as the first component of our new model by achieving the accuracy of 0.82, ranked as a second sentiment classification technique.

IV. Conclusion

In this work, we introduce a new model to detect misinformation in COVID-19 tweets based on the transfer learning approach by achieving nearly 90% accuracy. We use more than twenty different textual information resources based on the data provided by Twitter, NIH, CDC, Kaggle, George Town University to develop a model. A COVID-19 misinformation dataset is created and labeled by human annotators for this research. This dataset is used as ground truth to evaluate the final model performance. The deep context extraction of COVID-19 tweets, the computation time of training, memory usage, parallel computation ability, and classification accuracy is optimized in the new model for misinformation detection of COVID 19. Since we use the transfer learning approach, the model can classify unlabeled COVID-19 tweets on different online platforms based on the tweet's text. The new model and the new dataset can be used in public research on COVID-19 and Natural language research about the virus.

TABLE IX
THE NEW MODEL PERFORMANCE BASED ON DIFFERENT SENTIMENT SCORES GENERATED BY VARIUOS NLP MODELS

Sentiment Classification Models	f1-score	mcc
TextBlob	0.80	0.76
TF-IDF	0.73	0.65
BERT	0.80	0.75
RoBERTa	0.82	0.79
ALBERT	0.91	0.88

REFERENCES

[1] Abid Haleem, Mohd Javaid, "Effects of covid-19 pandemic in daily life." https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7147210/, 2020. Online; accessed 3 April 2020.

[2] R. A. Friedman, "Why humans are vulnerable to conspiracy theories." https://ps.psychiatryonline.org/doi/10.1176/appi.ps.202000348, 2020. Online; Retrived 10 October 2020.

[3] J. E. Oliver and T. Wood, "Medical Conspiracy Theories and Health Behaviors in the United States," *JAMA Internal Medicine*, vol. 174, pp. 817–818, 05 2014.

[4] D. Romer and K. H. Jamieson, "Conspiracy theories as barriers to controlling the spread of covid-19 in the u.s.," *Social Science & Medicine*, vol. 263, p. 113356, 2020.

[5] S. Laato, A. K. M. N. Islam, M. N. Islam, and E. Whelan, "What drives unverified information sharing and cyberchondria during the covid-19 pandemic?," *European Journal of Information Systems*, vol. 29, no. 3, pp. 288–305, 2020.

[6] C. P. Rodríguez, B. V. Carballido, G. Redondo-Sama, M. Guo, M. Ramis, and R. Flecha, "False news around covid-19 circulated less on sina weibo than on twitter. how to overcome false information?," *International and Multidisciplinary Journal of Social Sciences*, vol. 9, no. 2, pp. 107–128, 2020.

[7] G. Pennycook, J. McPhetres, Y. Zhang, J. G. Lu, and D. G. Rand, "Fighting covid-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention," *Psychological Science*, vol. 31, no. 7, pp. 770–780, 2020. PMID: 32603243.

[8] Z. Barua, S. Barua, S. Aktar, N. Kabir, and M. Li, "Effects of misinformation on covid-19 individual responses and recommendations for resilience of disastrous consequences of misinformation," *Progress in Disaster Science*, vol. 8, p. 100119, 2020.

[9] P. Cihan, "Fuzzy rule-based system for predicting daily case in covid-19 outbreak," in *2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, pp. 1–4, 2020.

[10] T. Tran, P. Rad, R. Valecha, and H. R. Rao, "Misinformation harms during crises: When the human and machine loops interact," in *2019 IEEE International Conference on Big Data (Big Data)*, pp. 4644–4646, 2019.

[11] M. Heidari and J. H. Jones, "Using bert to extract topic-independent sentiment features for social media bot detection," in *2020 11th IEEE Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON)*, pp. 0542–0547, 2020.

[12] S. Song, Y. Zhao, X. Song, and Q. Zhu, "The role of health literacy on credibility judgment of online health misinformation," in *2019 IEEE International Conference on Healthcare Informatics (ICHI)*, pp. 1–3, 2019.

[13] M. Angeline, Y. Safitri, and A. Luthfia, "Can the damage be undone? analyzing misinformation during covid-19 outbreak in indonesia," in *2020 International Conference on Information Management and Technology (ICIMTech)*, pp. 360–364, 2020.

[14] R. Sarkhel, M. Keymanesh, A. Nandi, and S. Parthasarathy, "Interpretable multi-headed attention for abstractive summarization at controllable lengths," in *Proceedings of the 28th International Conference on Computational Linguistics*, (Barcelona, Spain (Online)), pp. 6871–6882, International Committee on Computational Linguistics, Dec. 2020.

[15] A. Ghenai and Y. Mejova, "Catching zika fever: Application of crowdsourcing and machine learning for tracking health misinformation on twitter," in *2017 IEEE International Conference on Healthcare Informatics (ICHI)*, pp. 518–518, 2017.

[16] M. Heidari and S. Rafatirad, "Semantic convolutional neural network model for safe business investment by using bert," in *IEEE 2020 Seventh International Conference on Social Networks Analysis, Management and Security, SNAMS 2020*, 2020.

[17] A. Bridgman, E. Merkley, P. J. Loewen, T. Owen, D. Ruths, L. Teichmann, and O. Zhilin, "The causes and consequences of covid-19 misperceptions: understanding the role of news and social media," *The Harvard Kennedy School (HKS)*, 2020.

[18] "Covid-19 open research dataset challenge (cord-19)." https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge, 2020. Online; Retrived 21 September 2020.

[19] "Covid-19 twitter chatter dataset for scientific use." http://www.panacealab.org/covid19/, 2020.

[20] Şefika Şule Erçetin and N. B. Neyişci, *Social Network Analysis: A Brief Introduction to the Theory*. Springer, Cham, 2014.

[21] H. X. L. Ng and J. Y. Loke, "Analysing public opinion and misinformation in a covid-19 telegram group chat," *IEEE Internet Computing*, pp. 1–1, 2020.

[22] M. Heidari and S. Rafatirad, "Using transfer learning approach to implement convolutional neural network model to recommend airline tickets by using online reviews," in *2020 15th International Workshop on Semantic and Social Media Adaptation and Personalization (SMA*, pp. 1–6, 2020.

[23] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)* (J. Burstein, C. Doran, and T. Solorio, eds.), pp. 4171–4186, Association for Computational Linguistics, 2019.

[24] M. Heidari and S. Rafatirad, "Bidirectional transformer based on online text-based information to implement convolutional neural network model for secure business investment," in *IEEE 2020 International Symposium on Technology and Society (ISTAS20), ISTAS20 2020*, 2020.

[25] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," 2018.

[26] A. Radford, "Improving language understanding by generative pre-training," 2018.

[27] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP 2013, 18-21 October 2013, Grand Hyatt Seattle, Seattle, Washington, USA, A meeting of SIGDAT, a Special Interest Group of the ACL*, pp. 1631–1642, ACL, 2013.

[28] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized BERT pretraining approach," *CoRR*, vol. abs/1907.11692, 2019.

[29] Y. Zhu, R. Kiros, R. S. Zemel, R. Salakhutdinov, R. Urtasun, A. Torralba, and S. Fidler, "Aligning books and movies: Towards story-like visual explanations by watching movies and reading books," in *2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015*, pp. 19–27, IEEE Computer Society, 2015.

[30] S. Nagel, "Cc-news." https://commoncrawl.org/2016/10/news-dataset-available/, 2016.

[31] A. Gokaslan and V. Cohen, "Openwebtext corpus." https://skylion007.github.io/OpenWebTextCorpus/, 2016.

[32] T. H. Trinh and Q. V. Le, "A simple method for commonsense reasoning," *CoRR*, vol. abs/1806.02847, 2018.

[33] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "ALBERT: A lite BERT for self-supervised learning of language representations," in *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*, OpenReview.net, 2020.

[34] M. Heidari, J. H. J. Jones, and O. Uzuner, "Deep contextualized word embedding for text-based online user profiling to detect social bots on twitter," in *IEEE 2020 International Conference on Data Mining Workshops (ICDMW), ICDMW 2020*, 2020.

[35] "Amazonsagemakergroundtruth." https://aws.amazon.com/sagemaker/groundtruth/.

[36] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL* (A. Moschitti, B. Pang, and W. Daelemans, eds.), pp. 1532–1543, ACL, 2014.