2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI) | 978-1-6654-0898-1/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICTAI52525.2021.00141

Analysis of COVID-19 Misinformation in Social Media using Transfer Learning

Abhishek Dhankar
Department of Computing Science
University of Alberta
Edmonton, Alberta, Canada
dhankar@ualberta.ca

Hamman Samuel
Department of Computing Science
University of Alberta
Edmonton, Alberta, Canada
hwsamuel@ualberta.ca

Fahim Hassan
School of Public Health
University of Alberta
Edmonton, Alberta, Canada
fahim2@ualberta.ca

Nawshad Farruque
Department of Computing Science
University of Alberta
Edmonton, Alberta, Canada
nawshad@ualberta.ca

François Bolduc
Department of Pediatrics
University of Alberta
Edmonton, Alberta, Canada
fbolduc@ualberta.ca

Osmar Zaïane
Department of Computing Science
University of Alberta
Edmonton, Alberta, Canada
zaiane@ualberta.ca

***Abstract*—Most major events are often accompanied by misinformation on online Social Networking platforms. Due to its nature, the COVID-19 pandemic was bound to lead to an explosion of information online, much of it false or misleading. This information explosion, termed "infodemic" by the World Health Organization (WHO), has revealed the need for automatic fake news detection to help with the exponentially growing flow of unverified information. The objective of this study is to explore combinations of different supervised classification models trained on different general and domain-specific embeddings, and compare the effects of the iterations on the results. We also analyze the results to determine whether the differences in weighted F1-score performance metrics are statistically significant. Ultimately, we demonstrate that concatenation of general and context-specific embeddings improves performance. Our research shows promise for health misinformation detection and formulation of effective public health responses.**

***Keywords*-COVID-19, Misinformation, Social Media, Supervised Learning, Transfer Learning, Word Embeddings**

I. Introduction

The COVID-19 pandemic has played out as an infodemic, with misinformation, disinformation and rumours rapidly spreading on various facets of the disease such as origin, causes, symptoms, prevention, and treatments [1]. This has significantly hampered the global public health response. Social media is a popular way of communication, but uncertainties during the pandemic have caused proliferation of harmful health misinformation posts via platforms such as Reddit, Twitter, and Facebook, among others [2]. There are a number of topics fueling COVID-related misinformation, ranging from conspiracy theories, misreporting of morbidity and mortality, disease spread mechanisms, prevention methods, treatments and drugs, recovery experiences, and political controversies [3].

Although misinformation spreads both online and offline, the propagation and contagion of misinformation are more pronounced in social media platforms [4], [5]. Therefore, a critical understanding of the methods to detect misinformation in various social media platforms is a precursor to the design and implementation of effective health promotion policies [6]. One of the earlier attempts to detect health misinformation used Twitter data; Castillo et al. extracted multiple features from trending topic posts to classify the messages based on credibility [7]. Since then, there have been several interdisciplinary studies using social media (mostly using Twitter data) to understand the spread of misinformation [8]–[13], ranging from experiments on attitudes towards fake news [14], to public health policy frameworks [15], and conceptual theories in information and knowledge management [6]. In the field of Natural Language Processing (NLP), researchers have worked on building datasets related to misinformation, including representing with GloVe embeddings to find relevance between posts and misinformation [16]. Others have collected fact-checked articles covering a broad range of topics, including political and medical discussions [17], [18].

The objective of this study is to explore different supervised classification models, namely Support Vector Machine (SVM) and Multi-Layer Perceptron (MLP), trained on different embeddings. Various combinations of embeddings were used to determine whether these produced significant improvements in comparison to their constituent parts. Our research has practical implications during a pandemic when fact-checking activities are usually manual, and therefore, time-consuming, labour intensive and expensive. Insights from our study can help with development of automated systems which reduce the workload of manual fact-checking to clarify and debunk different types of misinformation.

978-1-6654-0898-1/21/$31.00 ©2021 IEEE
DOI 10.1109/ICTAI52525.2021.00141

II. Related Works

Our study uses NLP to analyze health misinformation specific to COVID-19. The challenges of more generic fake news detection from an NLP perspective can be categorized into four areas: fact-checking, rumor detection, stance detection, and sentiment analysis [19]. To facilitate the formulation of fake news as a supervised classification or regression task, various types of datasets have been used in literature, ranging from labelled short claims, e.g. PolitiFact and Snopes, to entire-article datasets where the whole article is either true or false. Labelled datasets for fake news detection in social networking services are limited. Various methods for general-purpose fake news detection have been utilized in literature, including machine learning models with and without neural networks, rhetorical approaches with Rhetorical Structure Theory (RST) to define the semantic role of text units and the overall coherence of a story, as well as Recognizing Textual Entailment (RTE) to recognize relationships between sentences. In relation to our research methodology, we use supervised machine learning for the COVID-19 misinformation detection task.

On the area of health misinformation, there is considerable work in literature, mainly covering vaccinations and infectious diseases. The findings from the related papers show notable prevalence of misinformation within social media posts [4]. Diving further into the specific topic of COVID-19 misinformation, one of the challenges has been inaccurate news coming from reputable sources on developing stories, such as efficacy of anti-inflammatory drugs [20]. At the same time, medical professionals have utilized social media more than ever before for sharing professional opinions and democratizing access to scientific data [20]–[22].

On the subject of COVID-19 misinformation, recent studies have also attempted to tackle this research challenge. Meng et al. fuse general embedding-based RoBERTa and COVID domain-specific embedding COVID-Twitter-BERT (CT-BERT) [23] using a simple MLP. The authors carried out the experiments on the aforementioned dataset to demonstrate that the combination of general and context specific embeddings marginally improves the performance of a classification model. However, they did not demonstrate that these improvements were statistically significant. Wani et al. [24] compare the performance of models based on general GloVe embeddings and domain specific fastText embeddings, which were trained on an non-annotated dataset of 179k COVID-related tweets posted by Gabriel Preda on Kaggle [25]. The word embeddings were not combined in any way by the authors. However, the context-specific fastText embedding did produce better results than the general GloVe embeddings. Here again no testing was done to determine whether those gains were statistically significant. We used the same COVID-related Kaggle dataset from [25] for training our own context-specific word embeddings.

From the sociological perspective, studies have shown that social media users share posts with misinformation mainly due to inattention to detail rather than any malicious intent [26]. Essentially, what people share on social media is not always what they believe. In the light of this, misinformation detection research can highlight trends that need public health interventions and repeated messaging [26]. Another factor to take into account is the sense of desperation that could be making people susceptible to misinformation. Research has shown that parents of ailing children are more likely to fall for online misinformation owing to desperation in finding treatments for chronic diseases like cancer [27]. In light of these factors, proper interventions on trending misinformation posts can help users to think more carefully about the accuracy of information they consume. Our work shines light on ways of improving misinformation detection on social media, thereby aiding effective public health responses. Additionally, once misinformation is identified, health professionals can also be enabled to engage with patients in social media to counter trending misinformation topics. As an example, pediatric infectious disease specialists have been proposed as a solution to social media misinformation about COVID-19 related to children and parents [28]. Ultimately, users consuming or spreading misinformation are usually not malicious, and once misinformation is detected, pro-science voices are imperative in social media.

III. Methodology

In this section, we discuss the dataset used, configurations for word embeddings, transfer learning approaches utilized, and the experiments conducted to find out if the different embeddings significantly improve the weighted F1 score.

A. Dataset and Preprocessing

The dataset used in our experiments was released at the CONSTRAINT 2021 workshop colocated with the 35th AAAI Conference on Artificial Intelligence [29]. Henceforth, we shall refer to this dataset as the CONSTRAINT_2021 dataset. This dataset contains 6420 and 2140 social media posts in the train and test sets respectively. These posts have been labelled as “real” and “fake”. The “real” posts were collected from official and verified Twitter handles, including government accounts, medical institutes, etc. [30]. The “fake” posts were collected from fact-checking websites such as Politifact, NewsChecker, and Boomlive. Fact-checked “fake” posts were collected regardless of the social media platforms that they were posted on, in contrast to the “real” social media posts which are entirely from Twitter. The combined CONSTRAINT_2021 dataset contains a total of 8,560 posts, of which 4,080 (47.7%) are “fake” posts, and 4,480 (52.3%) are “real” posts. All punctuation and standard stopwords were removed from the tweets. Thereafter, the tweet level representation was calculated and concatenations done as explained in the proceeding sections.

B. *Twitter Embeddings*

Word embeddings are vector representations of words where words with similar meaning share similar vector spaces. There are different ways of creating vector these representation. In our case we use Word2vec embeddings [31] for all our experiments. We used two word embeddings, namely General Twitter Embedding (GTE) and the Context-Specific Embedding (CSE), which we used to further derive all the other Twitter embeddings for our experiments.

1) General Twitter Embedding (GTE): For the general/universal embedding, we use a General Twitter Embedding (GTE) introduced in [32]. This embedding was trained on a corpus of 400 million tweets, and has a vocabulary size of 3 million words.

2) Context-Specific Embedding (CSE): To create embeddings specific to the COVID-19 context, we chose a corpus [25] posted on Kaggle containing 179,108 tweets spanning between 29th February, 2020 and 24th July, 2020. We used these tweets to create a Word2vec embedding of vector size 200. This embedding has a vocabulary size of 22,012 words, which is substantially lesser than the vocabulary of size of the GTE.

3) Tweet Level Vector Representations: This representation is created by taking the average across the word vector representations of all the words left in the tweet after pre-processing, provided said words are also in the vocabulary of the Word Embedding. This results in a single vector representation for each tweet, whose dimension will be equal to those of the individual words in the word embedding from which they were created. Therefore, SVM and MLP were have been used in our experiments, and not LSTMs, GRUs, and their transformer variants, which require sequential inputs.

C. *Transfer Learning*

We learn word embeddings on a small corpus [25] and then transfer the knowledge to the larger embedding space learned represented by a general corpus, thereby improving the representational power of the general embedding for a specific task. The task is creating representations of words in the context of the COVID-19 pandemic. Such an embeddding can take advantage of the large vocabulary size of a general embedding and the representational accuracy of the context specific word embedding. We carry out this transfer learning via two distinct ways. Firstly, using the method of transfer learning explained in [33], which we will be referring to as Augmentation Transfer Learning (ATL). Secondly, using concatenation of general and context specific embeddings to create new embeddings. This process will be called Concatenation Transfer Learning (CTL).

1) Augmentation Transfer Learning (ATL): We find out the common words between the vocabularies of GTE and CSE. The word embeddings of these common words are then used to train a simple neural network with ReLU activation function, which takes GTE word embeddings as input and the CSE word embeddings as the target output. This neural network, trained on common vocabulary word embeddings, can now be used to find the context-specific word embeddings of all the corresponding word embeddings in the GTE vocabulary, thereby creating a third representation called Augmented Twitter Embedding (ATE), which has the same vocabulary as GTE, but a vector size of CSE. This process by which the ATE is created is termed the Augmentation Transfer Leaning (ATL). The resulting ATL embedding does not have any duplicate words in its vocabulary.

2) Concatenation Transfer Learning (CTL): We use the General Twitter Embedding (GTE), Context Specific Embedding (CSE), and Augmented Twitter Embedding (ATE) to create concatenated embeddings. To create the concatenated tweet level vector representations, the tweet level vector representations explained in sub-section III-B3, are concatenated with each other. Two tweet level embeddings are created via this process of concatenation, one by concatenating GTE and ATE, called GTE+ATE, and the other by concatenating GTE and CSE, called GTE+CSE.

Overall, we use 5 different types of word embeddings to derive tweet level embeddings: (1) GTE, an off-the-shelf general word embedding for tweets, (2) CSE, a context specific word embedding for COVID-19 related tweets, (3) ATE, a context specific word embedding, but with a larger vocabulary than CSE, (4) GTE+CSE, a concatenation of the tweet level embeddings derived from GTE and CSE word embeddings, and (5) GTE+ATE, a concatenation of the tweet level embeddings derived from GTE and ATE word embeddings.

D. *Experiments*

Our objective with these experiments were two-fold: firstly, to determine if there was an improvement in the performance (determined by the weighted F1 score) of classification models when trained using embeddings created through Augmented Transfer Learning (namely ATE) and/or Concatenated Transfer Learning (namely GTE+CSE & GTE+ATE) vis-a-vis the embeddings from which the aforementioned embeddings were derived (namely GTE, CSE, and ATE). And secondly, to determine if the improvements thus produced were statistically significant.

In order to meet both these requirements, we combined the train and test sets of the CONSTRAINT_2021 dataset, and carried out 5x2 Cross Validation (CV) tests. The 5x2 CV involves carrying out 2-fold cross validation on the combined dataset across 5 iterations, with the dataset getting shuffled at every iteration. In each of the 5 iterations all the models are trained on one half of the dataset, and tested on the other half. In the same iteration, the training and testing halves are then swapped, and the models are trained on the erstwhile testing half and tested on the erstwhile training half. This produces a total of 10 test results for each model.

These 10 test results can then be used to test for whether a pair of models are statistically significantly different or not. This significance testing method, called the Combined 5x2 CV f-test, is done via the series of formulae specified in [34]: Let there be two models, namely A & B which need to be compared for statistical significance.

$$p^{(1)} = p_A^{(1)} - p_B^{(1)} \tag{1}$$

In Equation 1, $p_A^{(1)}$ is the vector of 5 weighted F1 scores which model A produced over the test set in the first split of the 2-CV, in each of the five iterations. Similarly for $p_B^{(1)}$. $p^{(1)}$ is the element-wise subtraction between $p_A^{(1)}$ & $p_B^{(1)}$.

$$p^{(2)} = p_A^{(2)} - p_B^{(2)} \tag{2}$$

In Equation 2, $p^{(2)}$ is similar to $p^{(1)}$, except the weighted F1 scores involved in this calculation were calculated in the second split of the 2-CV, in each of the 5 iterations.

$$\bar{p} = \frac{p^{(1)} + p^{(2)}}{2} \tag{3}$$

In Equation 3, $\bar{p}$ is the element-wise mean of the element-wise differences $p^{(1)}$ & $p^{(2)}$.

$$s^2 = \left(p^{(1)} - \bar{p}\right)^2 + \left(p^{(2)} - \bar{p}\right)^2 \tag{4}$$

In Equation 4, s^2 is the element-wise variance of the element-wise differences, $p^{(1)}$ & $p^{(2)}$. Finally, the f-statistic is calculated as follows:

$$f = \frac{\sum_{i=1}^{5}\sum_{j=1}^{2}\left(p_i^j\right)^2}{2\sum_{i=1}^{5} s_i^2} \tag{5}$$

The f-statistic is distributed with 10 and 5 degrees of freedom, for the numerator and denominator respectively. Those degrees of freedom along with the value of the f-statistic are used to determine the p-value for a pair of models. The two models in consideration are determined to be significantly different if the corresponding p-value < 0.05, i.e., we reject the null hypothesis that the two models are similar. This significance testing is carried out for every possible pairing of the models, and the conclusions are drawn accordingly.

E. Model Parameters

We used two classification models for our experiments, an SVM model, and an MLP model. SVM was a more traditional model, and the MLP was a stand in for Deep Learning models. The parameters for the SVM model are {`kernel:rbf, C:10, gamma:scale, random_state:8`}. $C = 10$ was configured based on hyperparameter tuning. The MLP contains two hidden layers of sizes 512 & 128 for the first and second layers respectively, the rest having the default values.

Model 1	Model 2	f-statistic	p-value	Significant?
GTE	ATE	81.785	0.000066	Yes
CSE	ATE	890.184	$<$ .00001	Yes
GTE	GTE+ATE	1.345	0.391441	No
ATE	GTE+ATE	64.005	0.000121	Yes
GTE	CSE	3.105	0.111458	No
GTE	GTE+CSE	16.561	0.00317	Yes
CSE	GTE+CSE	3.953	0.071234	No
GTE+ATE	GTE+CSE	18.385	0.00248	Yes

(a) SVM Results

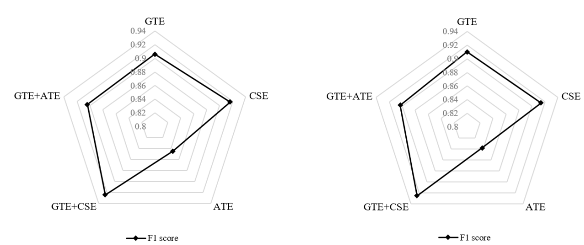
Model 1	Model 2	f-statistic	p-value	Significant?
GTE	ATE	114.177	0.000029	Yes
CSE	ATE	61.369	0.000135	Yes
GTE	GTE+ATE	1.099	0.488222	No
ATE	GTE+ATE	11.006	0.008146	Yes
GTE	CSE	2.055	0.220775	No
GTE	GTE+CSE	7.826	0.017418	Yes
CSE	GTE+CSE	8.107	0.016123	Yes
GTE+CSE	GTE+ATE	2.593	0.152293	No

(b) MLP Results

Table I: Results of 5x2-Fold CV for 2-Tailed Significance Testing with SVM and MLP Variants

The MLP was implemented using the Keras library, and the random state was set to 8 for reproducibility. We keep the parameters constant, regardless of the embedding being experimented on, because our objective is to determine whether changing the embedding alone can produce a significant change in the performance measure. If true, that change in the performance measure can be directly attributed to the change in embedding. Significance testing further confirms whether the change was statistically significant or just a fluke of random sampling. Each of the models will be subsequently referred to by the name of the embedding used to train the model. For instance, the SVM model trained on GTE will be simply referred to as GTE.

IV. Results and Discussion



(a) SVM Configurations (b) MLP Configurations

Figure 1: 5x2-fold CV Results for SVM and MLP

Figure 1 shows radar charts with the average of the 10 test results we got from the 5x2 CV experiment for different embeddings, for SVM and MLP models respectively. Tables Ib & Ia provide the relevant significance testing results to determine whether the differences in weighted F1-score performance metrics in Figure 1 are statistically significant.

Following is the analysis of the models trained on embeddings created through TL, as detailed in Section III-C. We discuss the most relevant and interesting results to appraise the overarching goal of our research towards COVID-19 misinformation detection.

A. ATE

Figure 1 shows that ATE based models perform worse that every other model for MLP & SVM, including models based on embeddings used to create ATE, namely GTE and CSE. Furthermore, Tables Ia & Ib shows that the performance difference is significant.

B. GTE+ATE

For SVM, Figure 1 shows that GTE+ATE performs as well as GTE, but it performs far better than ATE. Furthermore, Table Ia shows that the improvement over ATE is significant, but there is no significant difference between performances of GTE and GTE+ATE. It can be surmised that the improvement in the performance of GTE+ATE over ATE could be largely attributed to GTE. For MLP, GTE+ATE also performs far better than ATE, and as well as GTE. GTE+ATE's weighted F1-score is only 1% lesser than GTE's. While Table Ib shows that GTE+ATE's performance is significantly better than ATE's, there is no significant difference between GTE's and GTE+ATE's performances. We postulate that MLP was not able to take advantage of the concatenation of GTE and ATE. Again, improvement in GTE+ATE's performance can be attributed to GTE.

C. GTE+CSE

It can be clearly seen from Figure 1 that models based on GTE+CSE provide the best performance for both SVM and MLP across all the performance metrics. However, for SVM, Table Ia shows that while GTE+CSE performs significantly better than GTE, it does not show significant improvement over CSE. It can be surmised that for SVM, GTE+CSE's performance improvement is due to CSE. But, GTE+CSE has a vocabulary of size 3M, far larger than CSE's 22k, thereby lending GTE+CSE more generalizability because it can represent more words. Hence, GTE+CSE stands a better chance at outperforming CSE on datasets whose vocabulary may be very different from CSE. For MLP, Table Ib clearly shows that GTE+CSE's performance is significantly better than that of both GTE and CSE individually. While Figure 1 shows that on average GTE+CSE is better than GTE+ATE, but Tables Ib and Ia show that GTE+CSE performs significantly better than GTE+ATE for MLP, but the same performance is not observed for SVM. Ultimately, concatenation of general and context-specific embeddings significantly improves performance as shown by our analysis of the F1 score and the statistical significance tests. Such tests have not been carried out by comparable studies. This appraisal shows promise on using CTL for future research on health misinformation detection.

V. Conclusion

This study tackled the issue of health misinformation detection, and explored combinations of different supervised classification models trained on different general and domain-specific embeddings. We compared the effects of the iterations on the results and also analyze the results to determine whether the differences in weighted F1-score performance metrics were statistically significant. Ultimately, the concatenation approach of general and context-specific embeddings showed statistically significant improvement in performance. For future work, we plan to explore further the generalizability of Concatenation Transfer Learning with other datasets. Experiments using BERT and RoBERTA could be conducted in future works to further evaluate the findings in this short paper, given that transformers have been heavily utilized in recent research on similar NLP tasks.

References

[1] C. Pazzanese, "Battling the 'Pandemic of Misinformation'," 2020. Retrieved on 2020-05-08 from https://news.harvard.edu/gazette/story/2020/05/social-media-used-to-spread-create-covid-19-falsehoods/.

[2] K.-C. Yang, C. Torres-Lugo, and F. Menczer, "Prevalence of Low-Credibility Information on Twitter During the COVID-19 Outbreak," *arXiv 2004.14484*, 2020.

[3] E. Ferrara, "What Types of COVID-19 Conspiracies are Populated by Twitter Bots?," *First Monday*, vol. 25, May 2020.

[4] Y. Wang, M. McKee, A. Torbica, and D. Stuckler, "Systematic Literature Review on the Spread of Health-Related Misinformation on Social Media," *Social Science & Medicine*, vol. 240, p. 112552, 2019.

[5] D. M. Lazer, M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, *et al.*, "The Science of Fake News," *Science*, vol. 359, no. 6380, pp. 1094–1096, 2018.

[6] G. Eysenbach *et al.*, "How to Fight an Infodemic: The Four Pillars of Infodemic Management," *Journal of Medical Internet Research*, vol. 22, no. 6, p. e21820, 2020.

[7] C. Castillo, M. Mendoza, and B. Poblete, "Information Credibility on Twitter," in *Proceedings of the 20th International Conference on World Wide Web*, pp. 675–684, 2011.

[8] I. Hernández-García and T. Giménez-Júlvez, "Assessment of Health Information about COVID-19 Prevention on the Internet: Infodemiological Study," *JMIR Public Health and Surveillance*, vol. 6, no. 2, p. e18717, 2020.

[9] L. Singh, S. Bansal, L. Bode, C. Budak, G. Chi, K. Kawintiranon, C. Padden, R. Vanarsdall, E. Vraga, and Y. Wang, "A First Look at COVID-19 Information and Misinformation Sharing on Twitter," *arXiv preprint arXiv:2003.13907*, 2020.

[10] 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,” *Harvard Kennedy School Misinformation Review*, vol. 1, no. 3, 2020.

[11] V. Molter and R. DiResta, “Pandemics & Propaganda: How Chinese State Media Creates and Propagates CCP Coronavirus Narratives,” *Harvard Kennedy School Misinformation Review*, vol. 1, no. 3, 2020.

[12] S. B. Naeem, R. Bhatti, and A. Khan, “An Exploration of How Fake News is Taking Over Social Media and Putting Public Health at Risk,” *Health Information & Libraries Journal*, 2020.

[13] M. O. Lwin, J. Lu, A. Sheldenkar, P. J. Schulz, W. Shin, R. Gupta, and Y. Yang, “Global Sentiments Surrounding the COVID-19 Pandemic on Twitter: Analysis of Twitter Trends,” *JMIR Public Health and Surveillance*, vol. 6, no. 2, p. e19447, 2020.

[14] J. Agley and Y. Xiao, “Misinformation about COVID-19: Evidence for Differential Latent Profiles and a Strong Association with Trust in Science,” *BMC Public Health*, vol. 21, no. 1, pp. 1–12, 2021.

[15] V. Tangcharoensathien, N. Calleja, T. Nguyen, T. Purnat, M. D’Agostino, S. Garcia-Saiso, M. Landry, A. Rashidian, C. Hamilton, A. AbdAllah, *et al.*, “Framework for Managing the COVID-19 Infodemic: Methods and Results of an Online, Crowdsourced WHO Technical Consultation,” *Journal of Medical Internet Research*, vol. 22, no. 6, p. e19659, 2020.

[16] T. Hossain, R. L. Logan IV, A. Ugarte, Y. Matsubara, S. Singh, and S. Young, “Detecting COVID-19 Misinformation on Social Media,” *ACL 2020 Workshop*, 2020.

[17] G. K. Shahi and D. Nandini, “FakeCovid–A Multilingual Cross-domain Fact Check News Dataset for COVID-19,” *arXiv 2006.11343*, 2020.

[18] L. L. Wang, K. Lo, Y. Chandrasekhar, R. Reas, J. Yang, D. Eide, K. Funk, R. M. Kinney, Z. Liu, W. Merrill, P. Mooney, D. Murdick, D. Rishi, J. Sheehan, Z. Shen, B. Stilson, A. Wade, K. Wang, C. Wilhelm, B. Xie, D. Raymond, D. S. Weld, O. Etzioni, and S. Kohlmeier, “CORD-19: The Covid-19 Open Research Dataset,” *arXiv 2004.05125*, 2020.

[19] R. Oshikawa, J. Qian, and W. Y. Wang, “A Survey on Natural Language Processing for Fake News Detection,” *arXiv preprint arXiv:1811.00770*, 2018.

[20] D. Orso, N. Federici, R. Copetti, L. Vetrugno, and T. Bove, “Infodemic and the Spread of Fake News in the COVID-19 Era,” *European Journal of Emergency Medicine*, 2020.

[21] T. L. I. Diseases, “The COVID-19 Infodemic,” *The Lancet. Infectious Diseases*, vol. 20, no. 8, p. 875, 2020.

[22] A. Vaezi and S. H. Javanmard, “Infodemic and Risk Communication in the eEra of CoV-19,” *Advanced Biomedical Research*, vol. 9, 2020.

[23] X. Meng, W. Ren, and Y. Zhou, “Transformer-Based Language Model Fine-Tuning Methods 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, February 8, 2021, Revised Selected Papers*, p. 83, Springer Nature, 2021.

[24] A. Wani, I. Joshi, S. Khandve, V. Wagh, and R. Joshi, “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, February 8, 2021, Revised Selected Papers*, p. 153, Springer Nature, 2021.

[25] G. Preda, “Tweets with the Hashtag #covid19,” 2020. Retrieved on 2020-09-30 from https://www.kaggle.com/gpreda/covid19-tweets.

[26] G. Pennycook, T. D. Cannon, and D. G. Rand, “Prior Exposure Increases Perceived Accuracy of Fake News,” *Journal of Experimental Psychology*, vol. 147, no. 12, p. 1865, 2018.

[27] J. P. Guidry, C. A. Miller, A. J. Ksinan, J. M. Rohan, M. A. Winter, K. E. Carlyle, and B. F. Fuemmeler, “COVID-19–Related Misinformation among Parents of Patients with Pediatric Cancer,” *Emerging Infectious Diseases*, vol. 27, no. 2, p. 650, 2021.

[28] F. Scaggs Huang, P. Spearman, N. Baldwin, and J. K. Schaffzin, “Pediatric Infectious Disease Specialists: An Answer to Social Media Misinformation on Coronavirus Disease 2019,” *Journal of the Pediatric Infectious Diseases Society*, 12 2020.

[29] P. Patwa, M. Bhardwaj, V. Guptha, G. Kumari, S. Sharma, S. PYKL, A. Das, A. Ekbal, S. Akhtar, and T. Chakraborty, “Overview of CONSTRAINT 2021 Shared Tasks: Detecting English COVID-19 Fake News and Hindi Hostile Posts,” in *Proceedings of the First Workshop on Combating Online Hostile Posts in Regional Languages during Emergency Situation (CONSTRAINT)*, Springer, 2021.

[30] P. Patwa, S. Sharma, S. PYKL, V. Guptha, G. Kumari, M. S. Akhtar, A. Ekbal, A. Das, and T. Chakraborty, “Fighting an Infodemic: COVID-19 Fake News Dataset,” 2020.

[31] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient Estimation of Word Representations in Vector Space,” *arXiv preprint arXiv:1301.3781*, 2013.

[32] F. Godin, B. Vandersmissen, W. De Neve, and R. Van de Walle, “Multimedia Lab @ ACL W-NUT NER Shared Task: Named Entity Recognition for Twitter Microposts using Distributed Word Representations,” in *Proceedings of the Workshop on Noisy User-Generated Text*, pp. 146–153, 2015.

[33] N. Farruque, O. Zaiane, and R. Goebel, “Augmenting Semantic Representation of Depressive Language: From Forums to Microblogs,” in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 359–375, Springer, 2019.

[34] E. Alpaydm, “Combined $5\times$ 2 CV f test for Comparing Supervised Classification Learning Algorithms,” *Neural Computation*, vol. 11, no. 8, pp. 1885–1892, 1999.