Multi-Context Based Neural Approach for COVID-19 Fake-News Detection

Arkadipta De
ai20mtech14002@iith.ac.in
Department of Artificial Intelligence
Indian Institute of Technology Hyderabad
India

Maunendra Sankar Desarkar
maunendra@cse.iith.ac.in
Department of Computer Science and Engineering
Indian Institute of Technology Hyderabad
India

ABSTRACT

When the world is facing the COVID-19 pandemic, society is also fighting another battle to tackle misinformation. Due to the widespread effect of COVID 19 and increased usage of social media, fake news and rumors about COVID-19 are being spread rapidly. Identifying such misinformation is a challenging and active research problem. The lack of suitable datasets and external world knowledge contribute to the challenges associated with this task. In this paper, we propose **MiCNA**, a multi-context neural architecture to mitigate the problem of COVID-19 fake news detection. In the proposed model, we leverage the rich information of the three different pre-trained transformer-based models, i.e., *BERT*, *BERTweet* and *COVID-Twitter-BERT* to three different aspects of information (viz. general English language semantics, Tweet semantics, and information related to tweets on COVID 19) which together gives us a single multi-context representation. Our experiments provide evidence that the proposed model outperforms the existing baseline and the candidate models (i.e., three transformer architectures) and becomes a state-of-the-art model on the task of COVID-19 fake-news detection. We achieve new state-of-the-art performance on a benchmark COVID-19 fake-news dataset with **98.78**% accuracy on the validation dataset and **98.69**% accuracy on the test dataset.

CCS CONCEPTS

• **Computing methodologies** → **Language resources**; **Machine learning algorithms**.

KEYWORDS

COVID-19, Fake-News Detection, Transformers, BERT

ACM Reference Format:
Arkadipta De and Maunendra Sankar Desarkar. 2022. Multi-Context Based Neural Approach for COVID-19 Fake-News Detection. In *Companion Proceedings of the Web Conference 2022 (WWW '22 Companion), April 25–29, 2022, Virtual Event, Lyon, France.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3487553.3524662

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
WWW '22 Companion, April 25–29, 2022, Virtual Event, Lyon, France
© 2022 Association for Computing Machinery.
ACM ISBN 978-1-4503-9130-6/22/04…$15.00
https://doi.org/10.1145/3487553.3524662

1 INTRODUCTION

The usage of social media platforms such as Facebook, Twitter, etc., has increased tremendously, with many users worldwide engaging through these platforms every day. With the introduction of freedom of speech, every user can post, comment, and reply to anything related to any topic or domain. Fake news can be viewed as a piece of misinformation that is spread without any verification. Due to a lack of dedicated fact-checking/ monitoring organizations, misinformation on these social media platforms grows exponentially. The rapid spared of this COVID-19 related misinformation, including misreporting of ground realities, wrong measurements, unproven medications, have resulted in severe consequences like attacks on specific communities, mob lynching, hoarding of medicines and medical aids, etc.

When the whole world was under lockdown, social media like Facebook, Twitter, Whatsapp, Instagram, WeChat, etc., became significant sources of information for people. Research led by the Bruno Kessler Foundation in Italy showed that in March 2020, every day, there was an average of 46,000 new posts on Twitter linked to misleading information about the pandemic [10]. A recent survey [23] in the UK indicated that 46% of UK adults reported that they had been exposed to misleading information online about the crises. As pointed out in the report, 40% adults in the UK are "finding it hard to know what is true or false about the virus." Similarly, a study in the United States reported that 64% of US adults faced a great deal of confusion about the basic facts of current events due to the spread of fake news [4]. A recent study by World Health Organization (WHO) [24] shows that some of the most frequently spread myths during the COVID-19 pandemic are "Drinking alcohol protects you against COVID-19", "Spraying alcohol or chlorine all over your body kill the new coronavirus," "COVID-19 virus cannot be transmitted in areas with hot and humid climates", "Cold weather and snow kill the new coronavirus" and many more.

Fake news is spread primarily because of the absence of rigorous social media guidelines regarding the contents of posts. Also, the absence of fact-checking organizations in social media leads to people posting anything they want without even considering if the content is appropriate or not. The consequences of spreading fake news related to health conditions like COVID 19 are much more severe. In troubled times when people do not know how to take precautions and preventive measurements, fake news can lead to a disastrous outcome. In this paper we propose a neural architecture by leveraging three transformer based pre-trained models i.e., *BERT*, *BERTweet* and *COVID-Twitter-BERT*.

2 RELATED WORKS

Fake news detection is one of the most challenging, relevant, and essential problems in society, and there has been a rich line of work in fake-news detection. Below we discuss some of the most notable works in the line of fake-news detection in recent years.

2.1 Fake-news Detection in English

In an early approach, [1] presents a model to investigate the tension between information aggregation and the spread of misinformation. The authors showed that when the individuals exchange information, it can be modeled as both individuals adopting their pre-meeting beliefs. The cause and method of misinformation spread in social media are studied in [5]. The work also proposes necessary means to minimize the spread of misinformation. Characteristics of rumors in social media obtained by examining temporal, structural, and linguistic aspects of diffusion are presented in [16]. [31] presents a comprehensive review of detecting fake news on social media, including fake news characterizations on psychology, social theories, existing algorithms from a data mining perspective. A fake news dataset containing fake news and real news from different domains such as Entertainment, Business, Technology, Sports, Politics, etc., was released in [26]. [36] performed a comparative study on social media platforms and fake news detection methodology and strategies. [27] proposes a deep learning model based on Recurrent Neural Network to capture the temporal pattern of user activity on a given article. [15] proposes methods to combine information from different available sources and combine them to tackle the problem as Multi-source Multi-class Fake-news Detection (MMFD). In another work, [28] presents a hierarchical attention-based deep learning model for automatic multi-domain fake news detection.

2.2 Fake-news Detection in Other languages

There has been some active research in languages other than English. [2] released a novel dataset where they manually annotated 900 news articles (500 actual and 400 fake contents) from different domains (viz. Business, Health, Showbiz, Sports and Technology) in the Urdu language. Similarly, [13] provides an annotated dataset of around 50000 news articles in the Bengali language to mitigate the annotated resource scarcity problem in Bengali fake-news detection. Multiple models were developed for non-English languages as well. Other notable works on fake-news detection and misinformation are [8, 19, 22].

2.3 COVID-19 Fake-news Detection

There is also a rich line of work in fake news and misinformation spread detection in the field of COVID-19. Since the emergence of the pandemic COVID-19, researchers have been trying to use machine learning and deep learning-related techniques and frameworks to mitigate the problem of COVID 19 fake news and misinformation spread. [30] investigates a multilingual dataset consisting of 5182 fact-checked news articles on the topic of COVID-19, collecting news articles for about 90 different websites to speed up this research direction. [14] proposed a transformer-based architecture that uses both multilingual embeddings and 19 hand-crafted features for detecting fake news in Twitter posts. In the two-stage pipeline proposed in [34], the first stage retrieves the most relevant facts about a claim and the second stage verifies the truthfulness of the claim using textual entailment between the claims and the facts. [18] provides a multi-modal multilingual (MM) fake-news dataset related to COVID 19 and external social contexts. The dataset contains 3981 fake news content and 7192 trustworthy information from 5 different languages other than English.

Unlike previous approaches, we propose a multi context-based neural model for fake-news detection in COVID-19 related online posts in this paper. We use three different pre-trained transformers (pre-trained on different corpora) and leverage the average pooling of these representations in a neural architecture.

3 METHODOLOGY

3.1 BERT Architecture

The backbone of our proposed model is Transformer's Encoder-based pre-trained model BERT [7]. In this section, we briefly explain the BERT architecture and pre-training approach. BERT leverages the encoder component of the encoder-decoder architecture of the transformer, which consists of multiple layers, and each layer has components *viz.* self-attention, residual connections, feed-forward network, etc. It was trained on a large monolingual corpus based on a masked language objective (where we randomly mask 10-15% words and the model predicts masked words). BERT has proven as a state-of-the-art model in many Natural Language Understanding (NLU) tasks, including the General Language Understanding Evaluation (GLUE) benchmark [35], Squad, RACE, etc. An architecture diagram of BERT is shown in Figure 1

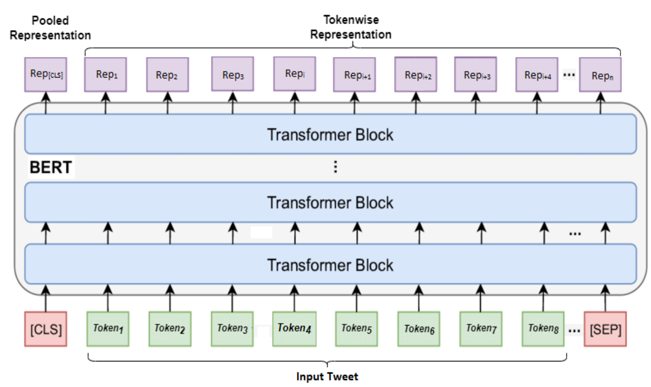


Figure 1: Basic BERT Model; Main Diagram Courtesy [11]

3.2 Multi-Context Neural Architecture

In this section, we discuss our proposed Multi-Context Neural Architecture (MiCNA) in detail. The architecture diagram of proposed model is given in Figure-2.

3.2.1 ***Input Representation:*** We capture the multi-context of each input token by utilizing three popular BERT based pre-trained models i.e *BERT*, *BERTweet* and *COVID-Twitter-BERT*. The BERT base [7] model has 12 layers and is pre-trained on English Wikipedia Corpus[1] (2,500M words) and the Book Corpus (800M words) [37].

[1]https://www.english-corpora.org/wiki/

It captures English language structure semantic and syntactic information. The *BERTweet* [21] model has the same architecture as the BERT base model, which is specifically trained on English Tweet corpus with masked language modeling objective. This model learns multiple unstructured latent features from tweets. To speed-up the research for preventing COVID 19 pandemic, *COVID-Twitter-BERT* (CT-BERT) model was proposed by [20]. Similar to *BERTweet*, authors trained BERT-base on a large corpus of COVID-19 related tweets with masked language modeling (MLM) objective. We feed the input post $P = p_1, p_2, p_3, \ldots p_k$ (where p_i is i^{th} token/sub-word of P) simultaneously into these three pre-trained models to obtain three different representations i.e., H_1, H_2 and H_3.

$$H_i = Model_i(P) \tag{1}$$

where $Model_i(i = 1, 2, 3)$ indicates one of the three BERT based pre-trained models. H_i is 1×768-dimensional $[CLS]$ token representation from i^{th} pre-trained model.

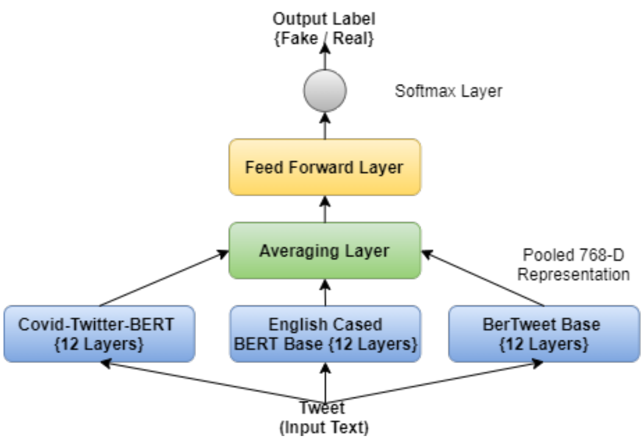


Figure 2: Multi-Context Transformer Based Architecture

3.2.2 ***Tokenization Layer:*** First, we perform tokenization of the input tweet before feeding it as an input to the model. We tokenize the tweet into constituent words and subwords. The significance of tokenizing a sentence into subwords is to handle the Out-Of-Vocabulary (OOV) word problems. There are two special tokens used in BERT: *[SEP]* and *[CLS]*. *[SEP]* token is used to denote the ending of a token sequence. *[CLS]* is used at the front of the sequence and corresponds to the "classification" token containing the learned compact representation of the entire input. Representation of this token is used in the subsequent task-specific layers. After appending these special tokens, we feed this tokenized sequence into transformer layers.

3.2.3 ***Averaging Pooling Layer:*** To avoid high input dimension for each post we have applied average pooling operation across three representations (H_1, H_2 and H_3) and obtained H_{avg}. It is a 768-dimension vector and fusion of latent features from three pre-trained models. We refer to it as a **multi-context** representation.

$$H_{avg} = \frac{1}{3}\sum_{i=1}^{3} H_i \tag{2}$$

3.2.4 ***Feed-forward Layer:*** The averaged representation H_{avg} is fed through a three layered Feed Forward Network *(FFN)* with GeLU [12] non-linear activation function to obtain representation F.

$$F = GeLU(FFN(H_{avg})) \tag{3}$$

3.2.5 ***Classification Layer:*** Finally, the hidden representation form FFN (i.e., F) is fed through the *Softmax* function to obtain the probability distribution of labels.

4 EXPERIMENTAL SETUP

4.1 Dataset

Sl.	Tweet	Label
1.	States reported 1121 deaths a small rise from last Tuesday Southern states reported 640 of those deaths.	Fake
2.	Clearly the Obama administration did not leave any kind of game plan for something like this.	Real

Table 1: Examples from COVID-19 FakeNews Dataset

Split	Real	Fake	Total
Training Set	3,360	3,060	6,420
Development Set	1,120	1,020	2,140
Testing Set	1,120	1,020	2,140
Total	5,600	5,100	10,700

Table 2: Statistics of Dataset

We use the COVID-19 fake-news dataset [25] for our experiments. This dataset is a collection of tweets related to COVID-19 news. The authors crawl tweets from verified Twitter handles using the Twitter API[2]. The considered Twitter handles were government accounts, medical institutes, news channels, sources from the World Health Organization (WHO), Centers for Disease Control and Prevention (CDC), Covid India Seva, Indian Council of Medical Research (ICMR), etc. It is expected that unauthorized or public accounts may include more fake information; they were also used on data collection. After data collection, authors utilize the fact-verification websites like PolitiFact[3], Snopes[4], Boomlive[5] etc. to check if a content is fake or not. Finally, human investigation is performed to verify and update the labels into two categories i.e., *Real* and *Fake*. It has total 10700 curated news records out of which 5600 and 5100 are labelled as *Real* and *Fake* respectively. The authors define *Real* and *Fake* labels as follows:

(1) **Real**: Tweets which are from verified sources and give useful information on COVID-19.
(2) **Fake**: Tweets, posts, articles which make claims and speculations about COVID-19 which are verified to be not true.

Table 1 and Table 2 tabulate example of dataset and distribution of data respectively.

[2]https://developer.twitter.com/en/docs/twitter-api
[3]https://www.politifact.com/
[4]https://www.snopes.com/
[5]https://www.boomlive.in/

Model Class	Model	Accuracy	Precision	Recall	Weighted F1 Score
Baselines [25]	Decision Tree	85.23	85.31	85.23	85.25
	Linear Regression	92.76	92.79	92.76	92.75
	Support Vector Machine	93.46	93.48	93.46	93.46
	Gradient Boosted Decision Tree	86.82	87.08	86.82	86.82
Current SOTA Models	La Diff ULMFit [3]	96.72	96.79	96.72	96.73
	Hybrid Transformer + Pseudo Labeling [17]	98.50	98.60	98.50	98.50
	ECOL + Prior Knowledge Injection [29]	97.57	97.66	97.57	97.57
	Ernie + Heuristic Decision [6]	97.60	97.60	97.60	97.60
MiCNA Components	BERT Base	95.98	96.02	95.96	95.98
	BERTweet	96.82	96.93	96.89	96.82
	COVID Twitter-BERT	97.66	97.70	97.63	97.66
Proposed	**MiCNA** (Dev Set)	**98.78**	**98.94**	**98.79**	**98.86**
	MiCNA (Test Set)	**98.69**	**98.83**	**98.57**	**98.64**

Table 3: Comparison of proposed model with baselines and current State-of-the-art models (SOTA); For baselines and SOTA, the results are reported on only test dataset

For additional experimentation, We perform several preprocessing techniques on the dataset. Pre-processing technique includes removal of non-alphanumeric characters (i.e., *@*, _, *$* etc.), newline and new paragraph characters. Additionally, we also experimented with removing stop-words and stemming. We tested model performance with and without preprocessed dataset.

4.2 Baseline and Evaluation Metrics

The baseline paper [25] uses term frequency–inverse document frequency (tf-idf) for representing each post. These representations are input for traditional machine learning classification algorithms. They leveraged following classification models:

(1) Support Vector Machines (SVM)
(2) Logistic Regression (LR)
(3) Decision Tree (DT)
(4) Gradient Boosted Decision Tree (GBDT)

To ensure that our proposed model is the best performing model, in addition to the above, we included three strong transformer-based baselines:

(1) BERT Base
(2) BERTweet
(3) COVID Twitter-BERT

These models are similar to the proposed MiCNA model. Unlike the MiCNA model, we used input representation from individual pre-trained models, not the pooled representation. We evaluated all the models performance on *accuracy*, *precision*, *recall* and *weighted average F1* metrics. Section 5 includes results and analysis of the experiments.

4.3 Implementation Details

We set the maximum input sequence length to 128, the warmup proportion to 0.15, the batch size to 28, and the number of epochs to 10. For all the models, the initial learning rate was set to $2e-5$. We use GeLU [12] as hidden activation function and *10%* Dropout [32] in the last layer of each pre-trained model model, *20%* Dropout to the Feed Forward Layer of MiCNA architecture. All other parameters of BERT Base[6], BERTweet[7] and COVID-Twitter-BERT[8] are not modified. We trained the model in an end-to-end manner. Hyperparameters are searched using a grid search algorithm.

5 RESULTS AND DISCUSSION

In this section, we discuss the experiments, results, and critical observations for the COVID-19 fake-news detection task. We compare our proposed architecture with the baseline models as defined by [25] and various models that achieve current state-of-the-art (SOTA) results. Table 3 includes results for all the baselines, current SOTA as well as the proposed architecture.

5.1 Comparison with Baselines

We conduct experiments with each of the three candidate transformer architectures, i.e., BERT-Base, BERTweet, and COVID-Twitter-BERT (CT-BERT), separately. We add a classifier layer on top of the three transformer architectures and finetune the models separately on the training dataset. For our proposed **MiCNA** model, we connect the three transformer components (c.f Figure 2) and train the model end to end. We conduct hyperparameter tuning using the development set and run inference using the best hyperparameters on the test dataset for all our experiments. The results of individual experiments on the test data are tabulated in Table 3.

It is observed that the three candidate transformer architectures (**viz.** BERT-Base, BerTweet, and COVID-Twitter-BERT) consistently outperformed all the existing baselines defined by [25] across all evaluation metrics. Finally, our proposed architecture **MiCNA** outperforms all the baselines as well as all the three candidate transformer architectures across all evaluation metrics. It is noteworthy that the evaluation scores of COVID-Twitter-BERT are close to the proposed model **MiCNA** which indicates the effect of the pre-trained model, which is trained on specifically COVID-19 dataset.

We also conducted experiments on our proposed model **MiCNA** with a different pre-processed version of the dataset (**viz.** removing punctuations, named entities). We observed that our model performed poorly, which gave evidence that every information (such

[6]https://github.com/google-research/bert
[7]https://github.com/VinAIResearch/BERTweet
[8]https://github.com/digitalepidemiologylab/covid-twitter-bert

as location, names especially named entities) is crucial in the online post due to its non-traditional sentence structure.

5.2 Comparison with Current State-of-the-art

We also compare our proposed architectures with different models that achieve current state-of-the-art (SOTA) on the COVID-19 Fake-News dataset. Different models employ different algorithms such as layered differentiated training approach for ULMFit [3], hybrid transformer architecture with pseudo labeling algorithm [17], injection of external knowledge with various transformer architectures [29] and transformer architectures coupled with heuristic decision making algorithm [6].

We observe from Table 3 that our proposed architecture outperforms these current SOTA models consistently. [17] uses an ensemble of transformers that achieves the closest accuracy score concerning our proposed architecture. Our proposed architecture also performs at per with the best performing model g2tmm [9][9] with a difference of **0.05**% with respect to weighted F1 Score.

6 ABLATION STUDY

We perform an ablation study on different components of our proposed **MiCNA** architecture to understand the importance of each component *viz.* BERT Base, BERTweet, and COVID Twitter-BERT. As opposed to the architecture of **MiCNA** (c.f Section 3.2) where we use averaging the representations of all the three components, in this study, we choose two of the three components and perform averaging. The results of this study are shown in Table 4.

Model Description	Accuracy	Precision	Recall	Weighted F1 Score
BERT + BERTweet	96.52	96.63	96.71	96.57
BERT + CT-BERT	97.83	97.87	97.63	97.79
BERTweet + CT-BERT	98.05	98.13	98.08	98.16

Table 4: Ablation Study for Component analysis of MiCNA architecture. CT-BERT: Covid-Twitter BERT

From Table 4, we observe that when we choose two of the three components of the MiCNA model, then the performance of the combined models surpasses the performance of the individual components as seen in Table 3, but they are not able to surpass the performance of final MiCNA model. This proves the importance of each component used in the MiCNA model where BERT understands the general English language constructs, BERTweet understands the general constructs structures of tweets, and COVID-Twitter BERT understands the scientific terms and information specific to COVID-19.

7 ANALYSIS

In this section, we further analyze the performances of the proposed model at a more granular level. More specifically, we try to understand answers to a few research questions mentioned:

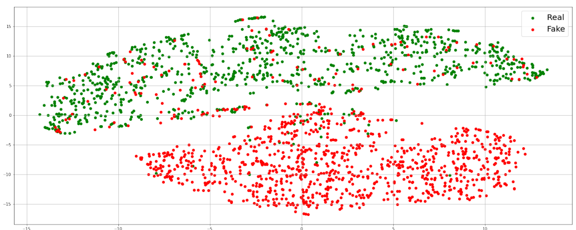
- **How does the proposed model perform for the individual classes?** We analyze the predictions obtained from MiCNA models on test data. The confusion matrix of the MiCNA model is included in Figure 5. We see that it is able to put most of the test examples in their corresponding correct classes. Moreover, with the number of mispredictions being roughly equal for real-predicted-as-fake and fake-predicted-as-real for this balanced dataset, it shows that the model is not biased towards any specific class.
- **Are the proposed multi-context model and the individual single-context models able to separate the examples from the different classes?** To understand whether the representations obtained by the individual models (BERT, BERTweet, and CT-BERT) as well as the proposed model (MiCNA) are able to separate the real and fake classes, we obtain the tSNE [33] plots of the representations. Figure 3 shows the representations in 2D. We see that there is a larger separation between the real and fake classes for the proposed multi-context (MiCNA) model, which indicates that it is beneficial to combine the strengths of the individual contexts to do better in the final prediction task.
- **How does the proposed multi-context model agree with the decisions of the individual single-context models?** We then try to see if the multi-context model indeed agrees with the predictions of the individual models and also whether it goes ahead with only one of the individual context representations. Table 5 shows that it agrees with the decisions of all the individual single-context-based models to a large extent. The table also shows that the proposed model does not prefer or go along with the decisions of *any single individual context*. However, it agrees slightly more with the CT-BERT model as CT-BERT has been pretrained on covid related corpus and can capture more context about COVID. We also see that the multi-context model disagrees with each of the individual models for quite a few examples. This again indicates that the multi-context model does not blindly follow any specific single-context model but indeed makes an attempt to further learn the classifier after obtaining the combined representation.

BERT - MiCNA	BERTweet - MiCNA	CT-BERT - MiCNA
1,826 (925/901)	1,943 (964/979)	1,962 (983/979)

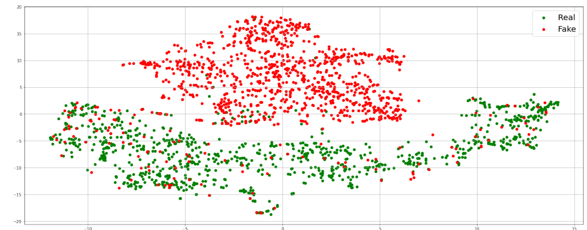
Table 5: Number of agreements of the multi-context model with individual single context models. Format: Total Agreement (Real Agreement/Fake Agreement). Number of examples: 2140, out of which 1120 are real and 1020 are fake.

- **Do Fake news surround specific types of named entities?** In this setting, we extract the top 12 most frequent Named Entities (NEs) from the predicted *Real* and predicted *Fake* tweets. Figures 4(a) and 4(b) include top-12 most frequent NEs for predicted fake and predicted real classes respectively. From Figure 4(a) we see that for predicted *Fake* tweets, most frequent named entities (NEs) are "**Trump**", "**India**", "**China**", "**Donald**", "**Wuhan**" etc. We observe from the Figure 4(b) that most frequent NEs for predicted *Real* tweets are "**India**", "**Lagos**", "**Auckland**" etc. We can say from the frequency of these NEs that predicted *Fake* tweets contain more political information rather than facts about COVID. On the other hand the predicted *Real* tweets contain more facts (e.g. such as occurrences of spread) related to COVID.
- **What are some of the posts that were predicted correctly by the multi-context model but were mispredicted by some of the single-context-based models?** We show some example tweets from the dataset where some of the MiCNA component

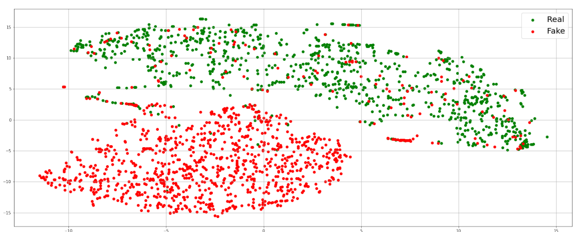
[9]The authors [9] only report Weighted F1 Score



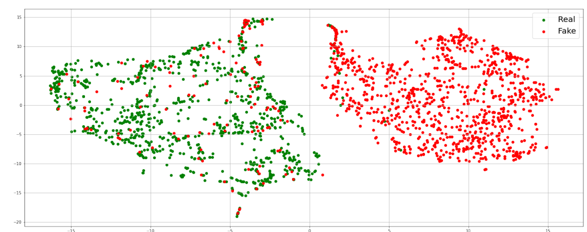
(a) Plot for BERT Representation



(b) Plot for BERTweet Representation

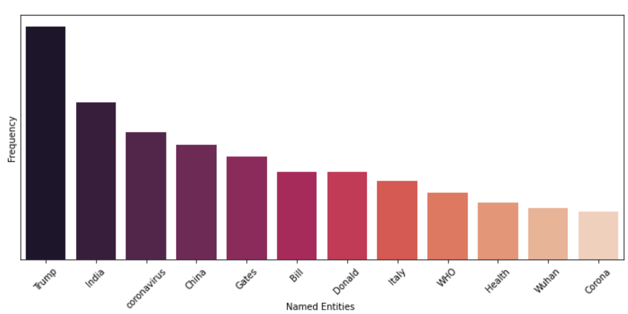


(c) Plot for CT-BERT Representation

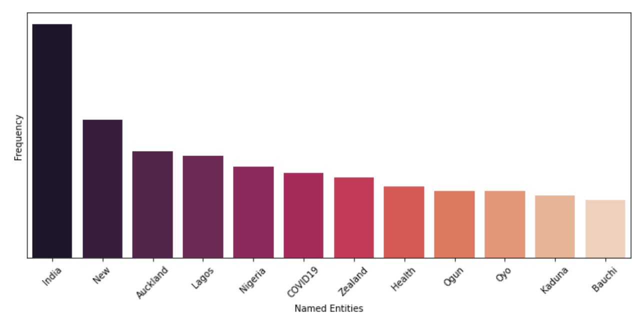


(d) Plot for MiCNA Representation

Figure 3: tSNE Embedding Plots for MiCNA and MiCNA Component Model Representations



(a) Most Frequent NEs for Predicted Fake Tweets



(b) Most Frequent NEs for Predicted Real Tweets

Figure 4: Named Entity Frequency Plot

Sl.	Tweet	Actual Label	Prediction by MiCNA Architecture	MiCNA Components that Misclassified
1.	Last note Washington DCs total test count fell by 22 presumably pulling out antibody tests #Coronavirus #USA	Real	Real	BERT, CT-BERT
2.	WHO reports record daily increase in global coronavirus cases up over 292000 #COVID #WHO	Fake	Fake	BERT, BERTweet
3.	Covid 19 is NOT killing people Weak immune systems and bad doctors are #Corona	Fake	Fake	BERT, BERTweet
4.	Russia has allegedly unleashed over 500 lions in order to ensure that people stay inside their houses #Russia #Covid19	Fake	Fake	BERTweet
5.	Thousands of doctors say hydroxychloroquine cures coronavirus #Cure #COVID	Fake	Fake	BERT, BERTweet

Table 6: Sample instances misclassified by at least one MiCNA component but correctly classified by MiCNA architecture

models (BERT-Base, BERTweet, CT-BERT) mispredict the classes of the tweets, but MiCNA correctly classifies them in Table 6.

- **Why does the proposed model perform better than the individual context-based models?** We understand that the semantic and syntactic structures of the tweets are quite different

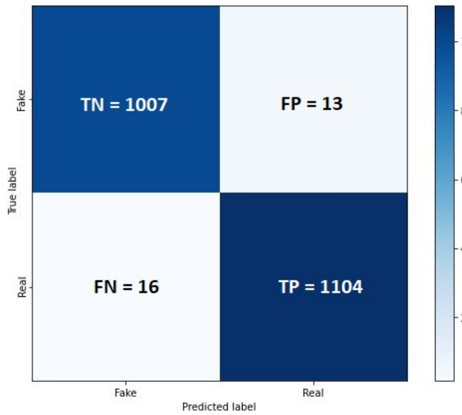


Figure 5: Confusion Matrix of MiCNA on Test set

from that of general English language literature texts. Moreover, to detect fake news about COVID-19, the model needs to know about the topic. Only English BERT-Base architecture is not enough to correctly classify these tweets. Hence we focus on an aggregate representation that comes from multiple contexts such as general English Language constructs, general Tweet constructs, and technical information related to COVID-19. This multi-context representation helps our model predict the class of the tweets more proficiently.

8 CONCLUSION AND FUTURE WORKS

In this paper, we tackle the problem of COVID-19 related fake-news detection. This problem is impactful because society must also take strong measures to prevent hoaxes or false information when the world is finding a cure for this pandemic disease. For this purpose, we propose an efficient Multi-Context based Transformer architecture that can efficiently classify the tweets related to COVID-19 as real and fake. In the future, we would like to use multimodal features for the fake news detection task.

REFERENCES

[1] Daron Acemoglu, Asuman Ozdaglar, and Ali ParandehGheibi. 2010. Spread of (mis)information in social networks. *Games and Economic Behavior* 70, 2 (2010), 194 – 227. https://doi.org/10.1016/j.geb.2010.01.005

[2] Maaz Amjad, Grigori Sidorov, Alisa Zhila, Helena Gómez-Adorno, Ilia Voronkov, and Alexander Gelbukh. 2020. "Bend the truth": Benchmark dataset for fake news detection in urdu language and its evaluation. *Journal of Intelligent & Fuzzy Systems* Preprint (2020), 1–13.

[3] Mohammed Azhan and Mohammad Kemal Ahmad. 2021. LaDiff ULMFiT: A Layer Differentiated training approach for ULMFiT. In *CONSTRAINT@AAAI*.

[4] Michael Barthel, Amy Mitchell, and Jesse Holcomb. 2016. Many Americans believe fake news is sowing confusion. *Pew Research Center* 15 (2016), 12.

[5] Ceren Budak, Divyakant Agrawal, and Amr El Abbadi. 2011. Limiting the Spread of Misinformation in Social Networks. In *Proceedings of the 20th International Conference on World Wide Web* (Hyderabad, India) *(WWW '11)*. Association for Computing Machinery, New York, NY, USA, 665–674. https://doi.org/10.1145/1963405.1963499

[6] Sourya Dipta Das, Ayan Basak, and Saikat Dutta. 2021. A Heuristic-driven Ensemble Framework for COVID-19 Fake News Detection. In *CONSTRAINT@AAAI*.

[7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. 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, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186. https://doi.org/10.18653/v1/N19-1423

[8] J. Gaglani, Y. Gandhi, S. Gogate, and A. Halbe. 2020. Unsupervised WhatsApp Fake News Detection using Semantic Search. In *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*. 285–289. https://doi.org/10.1109/ICICCS48265.2020.9120902

[9] Anna Glazkova, Maksim Glazkov, and T. Trifonov. 2021. g2tmn at Constraint@AAAI2021: Exploiting CT-BERT and Ensembling Learning for COVID-19 Fake News Detection. In *CONSTRAINT@AAAI*.

[10] Michael Goodyear. 2020. Fake News in the Time of COVID-19: Inherent Powers over Public Health. *Available at SSRN 3740639* (2020).

[11] Mutian He, Yangqiu Song, Kun Xu, and Yu Dong. 2020. On the Role of Conceptualization in Commonsense Knowledge Graph Construction. (03 2020).

[12] Dan Hendrycks and Kevin Gimpel. 2016. Gaussian Error Linear Units (GELUs). *arXiv: Learning* (2016).

[13] Md Zobaer Hossain, Md Ashraful Rahman, Md Saiful Islam, and Sudipta Kar. 2020. BanFakeNews: A Dataset for Detecting Fake News in Bangla. In *Proceedings of the 12th Language Resources and Evaluation Conference*. European Language Resources Association, Marseille, France, 2862–2871. https://www.aclweb.org/anthology/2020.lrec-1.349

[14] Debanjana Kar, Mohit Bhardwaj, Suranjana Samanta, and A. Azad. 2021. No Rumours Please! A Multi-Indic-Lingual Approach for COVID Fake-Tweet Detection. *2021 Grace Hopper Celebration India (GHCI)* (2021), 1–5.

[15] Hamid Karimi, Proteek Roy, Sari Saba-Sadiya, and Jiliang Tang. 2018. Multi-source multi-class fake news detection. In *Proceedings of the 27th International Conference on Computational Linguistics*. 1546–1557.

[16] Sejeong Kwon, Meeyoung Cha, K. Jung, Wei Chen, and Y. Wang. 2013. Prominent Features of Rumor Propagation in Online Social Media. *2013 IEEE 13th International Conference on Data Mining* (2013), 1103–1108.

[17] Xiangyang Li, Yu Xia, Xiang Long, Zheng Li, and Sujian Li. 2021. Exploring Text-transformers in AAAI 2021 Shared Task: COVID-19 Fake News Detection in English. In *CONSTRAINT@AAAI*.

[18] Yichuan Li, Bo han Jiang, Kai Shu, and Huan Liu. 2020. MM-COVID: A Multilingual and Multimodal Data Repository for Combating COVID-19 Disinformation. *ArXiv* abs/2011.04088 (2020).

[19] Jing Ma, Wei Gao, and Kam-Fai Wong. 2017. Detect Rumors in Microblog Posts Using Propagation Structure via Kernel Learning. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Vancouver, Canada, 708–717. https://doi.org/10.18653/v1/P17-1066

[20] M. Müller, M. Salathé, and P. Kummervold. 2020. COVID-Twitter-BERT: A Natural Language Processing Model to Analyse COVID-19 Content on Twitter. *ArXiv* abs/2005.07503 (2020).

[21] Dat Quoc Nguyen, Thanh Vu, and A. Nguyen. 2020. BERTweet: A pre-trained language model for English Tweets. In *EMNLP*.

[22] Van-Hoang Nguyen, Kazunari Sugiyama, Preslav Nakov, and Min-Yen Kan. 2020. FANG: Leveraging Social Context for Fake News Detection Using Graph Representation *(CIKM '20)*. Association for Computing Machinery, New York, NY, USA, 1165–1174. https://doi.org/10.1145/3340531.3412046

[23] Ofcom. 2020. Covid-19 news and information: consumption and attitudes.Results from week one of Ofcom's online survey. (2020). https://www.ofcom.org.uk/__data/assets/pdf_file/0031/193747/covid-19-news-consumption-week-one-findings.pdf

[24] World Health Organization et al. 2020. Coronavirus disease (COVID-19) advice for the public: myth busters. (2020).

[25] Parth Patwa, Shivam Sharma, Srinivas Pykl, Vineeth Guptha, Gitanjali Kumari, Md. Shad Akhtar, Asif Ekbal, Amitava Das, and Tanmoy Chakraborty. 2021. Fighting an Infodemic: COVID-19 Fake News Dataset. In *CONSTRAINT@AAAI*.

[26] Verónica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea. 2018. Automatic Detection of Fake News. In *Proceedings of the 27th International Conference on Computational Linguistics*. Association for Computational Linguistics, Santa Fe, New Mexico, USA, 3391–3401. https://www.aclweb.org/anthology/C18-1287

[27] Natali Ruchansky, Sungyong Seo, and Yan Liu. 2017. CSI: A Hybrid Deep Model for Fake News Detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (Singapore, Singapore) *(CIKM '17)*. Association for Computing Machinery, New York, NY, USA, 797–806. https://doi.org/10.1145/3132847.3132877

[28] Tanik Saikh, Arkadipta De, Asif Ekbal, and Pushpak Bhattacharyya. 2019. A Deep Learning Approach for Automatic Detection of Fake News. In *Proceedings of the 16th International Conference on Natural Language Processing*. 230–238.

[29] I. Baris Schlicht and Zeyd Boukhers. 2021. ECOL: Early Detection of COVID Lies Using Content, Prior Knowledge and Source Information. In *CONSTRAINT@AAAI*.

[30] Gautam Kishore Shahi and Durgesh Nandini. 2020. FakeCovid-A Multilingual Cross-domain Fact Check News Dataset for COVID-19. In *Workshop Proceedings of the 14th International AAAI Conference on Web and Social Media (ICWSM) 2020*. https://doi.org/10.36190/2020.14

[31] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter* 19, 1 (2017), 22–36.

[32] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* 15, 1 (Jan. 2014), 1929–1958.

[33] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research* 9, 86 (2008), 2579–2605. http://jmlr.org/papers/v9/vandermaaten08a.html

[34] Rutvik Vijjali, Prathyush Potluri, S. Kumar, and S. Teki. 2020. Two Stage Transformer Model for COVID-19 Fake News Detection and Fact Checking. *ArXiv* abs/2011.13253 (2020).

[35] Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R. Bowman. 2019. GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net. https://openreview.net/forum?id=rJ4km2R5t7

[36] Xinyi Zhou, Reza Zafarani, Kai Shu, and Huan Liu. 2019. Fake News: Fundamental Theories, Detection Strategies and Challenges. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining* (Melbourne VIC, Australia) *(WSDM '19)*. Association for Computing Machinery, New York, NY, USA, 836–837. https://doi.org/10.1145/3289600.3291382

[37] Yukun Zhu, Ryan Kiros, Rich Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. 2015. Aligning Books and Movies: Towards Story-Like Visual Explanations by Watching Movies and Reading Books. In *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV) (ICCV '15)*. IEEE Computer Society, USA, 19–27.