

Looking for COVID-19 misinformation in multilingual social media texts

Raj Ratn-Pranesh, Mehrdad Farokhnejad, Ambesh Shekhar, Genoveva Vargas-Solar

▶ **To cite this version:**

Raj Ratn-Pranesh, Mehrdad Farokhnejad, Ambesh Shekhar, Genoveva Vargas-Solar. Looking for COVID-19 misinformation in multilingual social media texts. 25th European Conference on Advances in Databases and Information Systems, University of Tartu, Aug 2021, Tartu, Estonia. 10.1007/978-3-030-85082-1_7 . hal-03314988

HAL Id: hal-03314988
https://hal.science/hal-03314988

Submitted on 5 Aug 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Looking for COVID-19 misinformation in multilingual social media texts

Raj Ratn Pranesh[1], Mehrdad Farokhnejad[2]
, Ambesh Shekhar[1], and Genoveva Vargas-Solar[4]

[1] Birla Institute of Technology, Mesra, India
raj.ratn18@gmail.com, ambesh.sinha@gmail.com
[2] Univ. Grenoble Alpes, CNRS, LIG, Grenoble, France
Mehrdad.Farokhnejad@univ-grenoble-alpes.fr
[3] CNRS, LIRIS-LAFMIA Lyon, France
genoveva.vargas-solar@liris.cnrs.fr

Abstract. This paper presents the Multilingual COVID-19 Analysis Method (CMTA) for detecting and observing the spread of misinformation about this disease within texts. CMTA proposes a data science (DS) pipeline that applies machine learning models for processing, classifying (Dense-CNN) and analyzing (MBERT) multilingual (micro)-texts. DS pipeline data preparation tasks extract features from multilingual textual data and categorize it into specific information classes (i.e., 'false', 'partly false', 'misleading'). The CMTA pipeline has been experimented with multilingual micro-texts (tweets), showing misinformation spread across different languages. To assess the performance of CMTA and put it in perspective, we performed a comparative analysis of CMTA with eight monolingual models used for detecting misinformation. The comparison shows that CMTA has surpassed various monolingual models and suggests that it can be used as a general method for detecting misinformation in multilingual micro-texts. CMTA experimental results show misinformation trends about COVID-19 in different languages during the first pandemic months.

Keywords: Misinformation · Multilingual Analysis · Micro-text analysis · COVID-19

1 Introduction

Since late 2019, the coronavirus disease COVID-19 has spread worldwide to more than 216 countries [15]. The COVID-19 pandemic has highlighted the extent to which the world's population is interconnected through the Internet and social media. Indeed, social media is a significant conduit where people share their response, thoughts, news, information related to COVID-19, with one in three individuals worldwide participating in social media, with two-thirds of people utilizing it on the Internet [16, 24]. Social media provides particularly fertile ground for the spread of information and misinformation [8]. It can even give direct access to content, which may intensify rumours and dubious information

[5]. For people with non-medical experience it is difficult to assess health information's authenticity. Misinformation may have intense implications for public opinion and behaviour, positively or negatively influencing the viewpoint of those who access it[3, 13]. Seeking accurate and valid information is the biggest challenge with Internet health information during the pandemic [7].

This paper proposes CMTA, a multilingual tweet analysis and information (misinformation) detection method for observing social media misinformation spread about the COVID-19 pandemic within communities with different languages. CMTA proposes a data science pipeline with tasks that rely on popular artificial intelligence models (e.g., multi-lingual BERT and CNN) to process texts and classifying them according to different misinformation classes ('false', 'partly false' and 'misleading'). The paper describes the experimental setting implemented for validating CMTA that uses datasets of COVID-19 related tweets. The paper presents an illustrative statistical representation of the findings insisting on the insights discovered in our study to discuss results. To assess the performance of CMTA, we compared CMTA with eight monolingual BERT models. The comparison shows that CMTA has surpassed various monolingual models and suggests that it can be used as a general method for detecting misinformation in multilingual micro-texts.

The remainder of the paper is organised as follows. Section 2 introduces works that have addressed misinformation detection about COVID-19 on social media datasets. Section 3 describes the general approach behind the method CMTA that we propose. It describes the experiment setting that we used for validating CMTA. Section 4 compares the performance of CMTA against mono-lingual analytics performed on the same dataset. It also discusses the study results about misinformation spread through micro-texts (i.e., tweets) across different languages. Finally, Section 5 concludes the paper and discusses future work.

2 Related work

The COVID-19 pandemic has resulted in studies investigating the various types of misinformation arising during the COVID-19 crisis [2, **?**,**?**,13]. Studies investigate a small subset of claims [22] or manually annotate Twitter data [13]. In [2] authors analyse different types of sources for looking for COVID-19 misinformation. Pennycook et al. [17] introduced an attention-based account of misinformation and observed that people tend to believe false claims about COVID-19 and share false claims when they do not think critically about the accuracy and veracity of the information. Kouzy et al. [13] annotated about 600 messages containing hashtags about COVID-19, they observed that about one-fourth of messages contain some form of misinformation, and about 17% contain some unverifiable information. With such misinformation overload, any decision making procedure based on misinformation has a high likelihood of severely impacting people's health [12]. The work in [11] examined the global spread of information related to crucial disinformation stories and "fake news" URLs during the early stages of the global pandemic on Twitter. Their study shows that news agencies,

government officials, and individual news reporters send messages that spread widely and play critical roles. Tweets citing URLs for "fake news" and reports of propaganda are more likely than news or government pages shared by regular users and bots.

The work in [21] focused on topic modelling and designed a dashboard to track Twitter's misinformation regarding the COVID-19 pandemic. The dashboard presents a summary of information derived from Twitter posts, including topics, sentiment, false and misleading information shared on social media related to COVID-19. Cinelli et al. [22] track (mis)-information flow across 2.7M tweets and compare it with infection rates.They noticed a major Spatio-temporal connection between information flow and new COVID-19 instances, and while there are discussions about myths and connections to low-quality information, their influence is less prominent than other themes specific to the crisis. To find and measure causal relationships between pandemic features (e.g. the number of infections and deaths) and Twitter behaviour and public sentiment, the work in [9] introduced the first example of a causal inference method. Their proposed approach has shown that they can efficiently collect epidemiological domain knowledge and identify factors that influence public interest and attention.

The discussion around the COVID-19 pandemic and the government policies was investigated in[14]. They used Twitter data in multiple languages from various countries and found common responses to the pandemic and how they differ across time using text mining. Moreover, they presented insights as to how information and misinformation were transmitted via Twitter. Similarly, to demonstrate the epidemiological effect of COVID-19 on press publications in Bogota, Colombia, [19] used text mining on Twitter data. They intuitively note a strong correlation between the number of tweets and the number of infected people in the area.

Most of the works described above focus on analyzing tweets related to a single language such as English. Our work has designed a single model leveraging multilingual BERT for the analysis of tweets in multiple languages. Furthermore, we used a large data set to train and analyze the tweets. We aim to provide a system that will not be restricted to any language for analyzing social media data.

3 The CMTA Method

Both misinformation[4] and disinformation[5], according to the Oxford English Dictionary, are false or misleading information. Misinformation refers to information that is accidentally false and spread without the intent to hurt, whereas disinformation refers to false information that is intentionally produced and shared to cause hurt [10]. Claims do not have to be entirely truthful or incorrect; they can contain a small amount of false or inaccurate information[20]. This work uses the

[4] https://www.oed.com/view/Entry/119699?redirectedFrom=misinformation
[5] https://www.oed.com/view/Entry/54579?redirectedFrom=disinformation

general notion of misinformation and makes no distinction between misinformation and disinformation as it is practically difficult to determine one's intention computationally.

The data science pipeline phases proposed by CMTA are: tokenizing, text features extraction, linear transformation, and classification. The first phases (tokenizing, text feature extraction, linear transformation) correspond to a substantial data-preparation process intended to build a multi-lingual vectorized representation of texts. The objective is to achieve a numerical pivot representation of texts agnostic of the language. CMTA classification task uses a dense layer and leads to a trained network model that can be used to classify micro-texts (e.g. tweets) into three misinformation classes: 'false', 'partly false' and 'misleading'.

Text tokenization Given a multilingual textual dataset consisting of sentences, CMTA uses the BERT multilingual tokeniser to generate tokens that BERT's embedding layer will further process. CMTA uses MBERT[6] to extract contextual features, namely word and sentence embedding vectors, from text data [7]. In the subsequent CMTA phases that use NLP models, these vectors are used as feature inputs with several advantages. (M)BERT embeddings are word representations that are dynamically informed by the words around them, meaning that the same word's embeddings will change in (M)BERT depending on its related words within two different sentences.

For the non-expert reader, the tokenization process is based on a WordPiece model. It greedily creates a fixed-size vocabulary of individual characters, subwords, and words that best fit a language data (e.g. English) [8]. Each token in a tokenized text must be associated with the sentence's index: sentence 0 (a series of 0s) or sentence 1 (a series of 1s). After breaking the text into tokens, a sentence must be converted from a list of strings to a list of vocabulary indices. The tokenisation result is used as input to apply BERT that produces two outputs, one pooled output with contextual embeddings and hidden-states of each layer. The complete set of hidden states for this model are stored in a structure containing four elements: the layer number (13 layers) [9], the batch number (number of sentences submitted to the model), the word / token number in a sentence, the hidden unit/feature number (768 features) [10].

[6] https://github.com/google-research/bert/blob/master/multilingual.md

[7] Embeddings are helpful for keyword/search expansion, semantic search and information retrieval. They help accurately retrieve results matching a keyword query intent and contextual meaning, even in the absence of keyword or phrase overlap.

[8] This vocabulary contains whole words, subwords occurring at the front of a word or in isolation (e.g., "em" as in the word "embeddings" is assigned the same vector as the standalone sequence of characters "em" as in "go get em"), subwords not at the front of a word, which are preceded by '##' to denote this case, and individual characters [25]

[9] It is 13 because the first element is the input embeddings, the rest is the outputs of each of BERT's 12 layers.

[10] That is 219,648 unique values to represent our one sentence!

In the case of CMTA, the tokenisation is more complex because it is done for sentences written in different languages. Therefore, it relies on the MBERT model that has been trained for this purpose.

Feature Extraction Phase is intended to exploit the information of hidden-layers produced due to applying BERT to the tokenisation phase result. The objective is to get individual vectors for each token and convert them into a single vector representation of the whole sentence. For each token of our input, we have 13 separate vectors, each of length 768. Thus, to get the individual vectors, it is necessary to combine some of the layer vectors. The challenge is to determine which layer or combination of layers provides the best representation.

Linear convolution The hidden states from the 12th layer are processed in this phase, applying linear convolution and pooling to get correlation among tokens. We apply a three-layer 1D convolution over the hidden states with consecutive pooling layers. The final convolutional layer's output is passed through a global average pooling layer to get a final sentence representation. This representation holds the relation between contextual embeddings of individual tokens in the sentence.

Classification A linear layer is connected to the model in the end for the CMTA classification task. This classification layer outputs a Softmax value of vector, depending on the output, the index of the highest value in the vector represents the label for the given sequence: 'false', 'partly false' and 'misleading'.

3.1 Experiment

To validate CMTA, we designed experiments on Google Colab with 64 GB RAM and 12 GB GPU. For implementing the method, we calibrated pre-trained models provided by hugging face[11]. For our experimental setting, we extracted annotated misinformation data from multiple publicly available open databases. We also collected many multilingual tweets consisting of over 2 million tweets belonging to eight different languages.

Misinformation datasets We collected data from an online fact-checker website called Poynter [18]. Poynter has a specific COVID-19 related misinformation detection program named 'CoronaVirusFacts/DatosCoronaVirus Alliance Database[12]'. This database contains thousands of labelled social media information such as news, posts, claims, articles about COVID-19, which were manually verified and annotated by human volunteers (fact-checkers) from all around the globe. The database gathers all the misinformation related to COVID-19 cure, detection, the effect on animals, foods, travel, government policies, crime, lockdown. The misinformation dataset is available in 2 languages- 'English' and 'Spanish'.

[11] https://huggingface.co/
[12] https://www.poynter.org/covid-19-poynter-resources/

We crawled through the content of two websites using Beautifulsoup[13], a Python library for scraping information from web pages. We scraped 8471 English language false news/information belonging to nine classes, namely, 'False', 'Partially false', 'Misleading', 'No evidence', 'Four Pinocchios', 'Incorrect', 'Three Pinocchios', 'Two Pinocchios' and 'Mostly False'. We gathered the article's title, its content, and the fact checker's misinformation-type label for each article.

For Spanish[14], we collected 531 misinformation articles. The collected data contains the misinformation published on social media platforms such as Facebook, Twitter, Whatsapp, YouTube. Posts were mostly related to political-biased news, scientifically dubious information and conspiracy theories, misleading news and rumours about COVID-19. We also used a human-annotated fact-checked tweet dataset [1] available at a public repository[15]. The dataset contained true and false labelled tweets in English and Arabic language. We used only false labelled tweets consisting of 500 English. We compiled a total of 9,502 micro-articles distributed across 9 misinformation classes shown in Table 1.

Table 1: Collected Misinformation Data set.

Classes	**Number of tweets**
False [18] (English)	2,869
Partially False (English)	2,765
Misleading (English)	2,837
False (Spanish)	191
Partially False (Spanish)	161
Misleading (Spanish)	179
False [1] (English)	500
Total	9,502

Dataset Pre-processing The datasets contained noise such as emojis, symbols, numeric values, hyperlinks to websites, and username mentions that were needed to be removed. Since our dataset was multilingual, we had to be very careful while preprocessing as we did not want to lose any valuable information. To preprocesses the training and inference datasets, we used simple regular expressions to remove URLs, special characters or symbols, blank rows, re-tweets, user mentions. We did not remove the hashtags from the data as hashtags might contain helpful information. For example, in the sentence- 'Wear mask to protect yourself from #COVID-19 #corona', only the symbol '#' was removed. We removed stop words using NLTK[16]. NLTK supports multiple languages except for few

[13] https://pypi.org/project/beautifulsoup4/
[14] https://chequeado.com/latamcoronavirus/
[15] https://github.com/firojalam/COVID-19-tweets-for-check-worthiness
[16] NLTK https://www.nltk.org/ is a Python library for natural language processing.

languages, such as Hindi and Thai. For preprocessing the Hindi dataset, we used CLTK (Classical Language Toolkit) [17]. For removing Thai stop words from Thai tweets, we used PyThaiNLP [23]. The emojis were removed using their Unicode.

Attribute engineering of the training dataset Table 2 gives an overview of the training dataset and showcase some misinformation articles. Column 2 shows the original label assigned by the fact-checker, column 3 gives a misinformation example associated with the label present in column 2, and column 4 provides reasoning given by the fact-checker behind assigning a particular label (column 2) to the misinformation (column 3). For example, if we look at the entry number '3' in the table 2, the misinformation is about the adverse effect of 5G radiation over the COVID-19 patients. This entry was labelled 'Incorrect' by the fact-checker. After analysing the fact-checker rating and the explanation given, we labelled it as 'False' misinformation. Entry number '5' talks about the COVID-19 test cost. The explanation given by the fact-checker is valid as it is not sure if there is any fee in the USA for the COVID-19 test or not. So because of the lack of evidence and uncertainty, we labelled it as 'Partially false'. Entry number '7' in the table talks about a video showing COVID-19 corpus dumping in the sea. Based on the explanation, the video was coupled with the wrong information to mislead the audience. So it was labelled as 'Misleading' misinformation.

The collected data is unevenly distributed across nine classes: 'No evidence', 'Four Pinocchios[18]', 'Incorrect', 'Three Pinocchios[19]', 'Two Pinocchios[20]' and 'Mostly False' (the smallest group). Most collected articles were labelled either as 'False', 'Partially false' and 'Misleading'. We performed an attribute engineering phase for preparing the dataset. We produced a uniformly distributed dataset reorganised the initial dataset as follows. The classes 'Four Pinocchios' and 'Incorrect' were merged with the class 'False'. The classes 'Three Pinocchios' and 'Two Pinocchios' were merged into the class 'Partially false'. The classes 'No evidence' and 'Mostly False' were merged with the class 'Misleading'. Finally, column 1 (see table 2) corresponds to the label assigned during the attribute engineering phase.

Inference Dataset For building an inference dataset to be used to test the CMTA trained model for analysing the misinformation spread across all over the social media platforms in multiple languages, We collected around 2,137,106 multilingual tweets. The tweets were expressed in eight major languages, namely- 'English', 'Spanish', 'Indonesian', 'French', 'Japanese', 'Thai', 'Hindi' and 'German'. Therefore, we used a dataset of tweets IDs associated with the novel coronavirus COVID-19 [4]. Starting on January 28, 2020, the current dataset contains 212,978,935 tweets divided into groups based on their publishing month. The dataset was collected using multilingual COVID-19 related keywords and con-

[17] https://docs.cltk.org/en/latest/index.html
[18] 90%-95% changes of it being false
[19] 70%-75% changes of it being false
[20] 50%-55% changes of it being false

Table 2: Misinformation Dataset

Our Rating	IFCN(Poynter) Rating	Misinformation	Explanation
False	False	The border between France and Belgium will be closed.	French and Belgian authorities denied it.
	Four pinocchios	Trump's effort to blame Obama for sluggish coronavirus testing.	There was no "Obama rule," just draft guidance that never took effect and was withdrawn before President Trump took office.
	Inaccurate	Elisa Granato, the first volunteer in the first Europe human tria of a COVID-19 vaccine, has died.	Elisa Granato, the first volunteer in the first Europe human trial of a COVID-19 vaccine, has died.
Partially False	Partially False	Media shows a Florida beach full of people while it's empty.	The different videos were not shot at the same time. The beaches are empty when they are closed.
	Two Pinocchios	The bill for a coronavirus test in the US is $3.000	The CDC is not making people pay the test by now.
	Partly False	Salty and sour foods cause the "body of the COVID-19 virus" to explode and dissolve.	"Consuming fruit juices or gargling with warm water and salt does not protect or kill COVID-19," the World Health Organization Philippines told VERA Files.
Misleading	Misleading	A clip from Mexico depicts the dumping of coronavirus patients corpses into the sea.	Misbar's investigation of the video revealed that it does not depict the dumping of coronavirus patients corpses in Mexico, but rather paratroopers landing from a Russian MI 26 helicopter.
	No Evidence	Media uses photos of puppets on patient stretchers to scare the public.	There is no evidence that any media outlet used this photo for their reporting about COVID-19. Its origin is unclear, maybe it was shot in Mexico and shows a medical training session.
	Mostly False	Coronavirus does not affect people with 'O+' blood type.	The post claiming coronavirus does not affect people with 'O+' blood type is misleading.

tained tweets in more than 30 languages. We used tweepy[21] which is a Python module for accessing Twitter API. We decided to retrieve the tweets using the tweet IDs published in the past five months (February, March, April, May and June) for our analysis. Table 3 shows the total number of collected tweets. The distribution of tweets across eight languages corresponds to most English items, almost 1 and 1/8 of the whole data set, then Spanish (1/4 of the total number of tweets) and the rest for French, Japanese, Indonesian, Thai, and Hindi.

3.2 Model Setup and Training

Training Setting For training our model, we divided the data into training, validation and testing datasets in the ratio of 80%/10%10% respectively. The final count for the train, validation and test dataset was 7,602, 950, 950. We fine-tuned the Sequence Classifier from HuggingFace based on the parameters specified in [6]. Thus, we set a batch size of 32, learning rate 1e-4, with Adam Weight Decay as the optimizer. We run the model for training for 10 epochs. Then, we save the model weights of the transformer, helpful for further training.

Hyperparameters' Setting Table 4 lists every hyperparameter for training and testing our model. All the calculations and selection of hyperparameters were made based on tests and the model's best output. After performing several

[21] Python module is available at http://www.tweepy.orghttp://www.tweepy.org

Table 3: Language-wise Dataset Distribution

Language	**ISO**	**Number of tweets**
English	en	1,472,448
Spanish	es	353,294
Indonesian	in	80,764
French	fr	71,722
Japanese	ja	71,418
Thai	th	36,824
Hindi	hi	27,320
German	de	23,316
Sum		2137106

iterations on distinct sets of hyper-parameters based on the model's performance analysis, we adopted the one showing promising results on our dataset.

Table 4: Hyper-parameters for training

Parameters	**Value**
Pool Size of Average Pooling	8
Pool Size of Max Pooling	8
Dropout Probability	0.36
Number of Dense layers	4
Text Length	128
Batch Size	32
Epochs	10
Optimizer	Adam
Learning Rate	1×10^{-4}

3.3 Experiment and Results

We experimented with the multilingual data with their respective linguistic-based BERT models. We set the model with the same training parameters as the CMTA model and preprocessed the data as stated previously. Each monolingual model was fine-tuned for 10 epochs with a batch size of 32, and it was applied to the classification dataset of their respective language. Our model achieved an accuracy(%) of **82.17** (see figure 1)and F_1 score (%) of **82.54** on the test dataset. The precision and recall reported by the model were **82.07** and **82.30** respectively.

Table 5 shows the model's prediction over few examples from the test dataset along with their actual label. As we have shown in the table, in the entry numbers '1', '2', '3' and '4' our model could predict the correct label. However, in the case of entry number '5', the label predicted by our model was 'False', whereas the

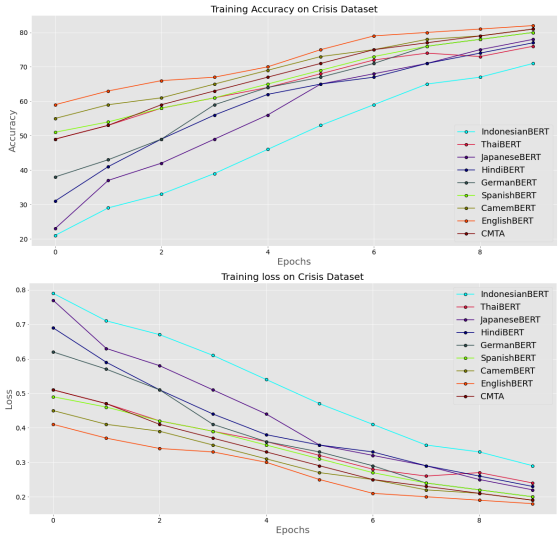


Fig. 1: Training Accuracy(Upper) and Training loss(Lower)

actual label is 'Misleading'. If we would look at the misinformation at the entry number '5', which is a Spanish text- 'El medicamento contra piojos sirve como tratamiento contra Covid-19.' and the corresponding English translation would be- ". This misinformation claims about a COVID-19 medicine, and since this could be 'false' and 'misleading' misinformation at the same time, our model predicted it as 'false' misinformation rather than 'misleading'.

Table 5: Misinformation data examples along with model's prediction and actual label

Test Data	**Actual Label**	**Prediction**	**Accuracy(✓/✗)**
Dr. Megha Vyas from India died due to COVID-19 while treating COVID patients.	False	False	✓
El plátano bloquea "la entrada celular del COVID-19"	False	False	✓
Asymptomatic people are very rarely contagious, said the WHO.	Partially False	Partially False	✓
Patanjali Coronil drops can help cure coronavirus.	Misleading	Misleading	✓
El medicamento contra piojos sirve como tratamiento contra Covid-19.	Misleading	False	✗

4 Results Assessment

We used two strategies for assessing CMTA results according to the research questions we wanted to prove. The first research question was to determine (Q_1) whether it was possible to develop a method that could provide a general multi-lingual classification pipeline? The second question was (Q_2) whether it is possible to build conclusions about how misinformation on COVID-19 spreads in different language speaking communities by analysing micro-texts published in social media. For Q_1, we conducted a comparative study of CMTA with different independent mono-lingual misinformation classifiers. Thereby, CMTA's classification performance for a given set of micro-texts written in a given language was compared against a classification model targeting only that language. For Q_2 we analysed and plotted CMTA's results, and we proposed intuitive arguments according to our observations.

4.1 CMTA vs Monolingual Classification

We conducted a comparative performance study of various monolingual BERT models concerning our proposed multilingual CMTA model for comparing their performance for the misinformation detection task. We investigated eight monolingual BERT model[22], namely, 'English', 'Spanish', 'French', 'Germann', 'Japanese', 'Hindi' 'Thai[23]' and 'Indonesian'. We used the same 9,502 tweets distributed across three misinformation classes for training the monolingual models. Our dataset consisted of English and Spanish tweets; therefore, we translated the tweets into eight languages to train each of the eight monolingual models. We used Google Translator API[24] for converting the tweets into a particular language. Results are shown in Table 6. Based on the experiment results, we strongly suggest that the multilingual CMTA model could generalize smoothly on the dataset because its performance was equivalent to the monolingual models.

4.2 Multilingual Misinformation Analysis

We provide a detailed analysis of misinformation distribution across multilingual tweets. This analysis responds to the initial question: *how is misinformation about COVID-19 spread in communities speaking different languages.* Our survey studied and analyzed the distribution of COVID-19 misinformation across eight significant languages (i.e. 'English', 'Spanish', 'Indonesian', 'French', 'Japanese', 'Thai', 'Hindi' and 'German') for five months (i.e. February, March, April, May and June).

We used our trained, multilingual model, CMTA, to predict and categorize the misinformation type present in tweets. We conducted our sequential misinformation analysis on a collection of over 2 million multilingual tweets. Figure 2 shows the month-wise distribution of misinformation types for each language.

[22] Pretrained model available at https://huggingface.co/models
[23] ThaiBERT is available at https://github.com/ThAIKeras/bert
[24] Please refer https://cloud.google.com/translate/docs

Table 6: Precision, Recall and f-score of CMTA model

Metrics / Models	Precision	Recall	F1-score
EnglishBERT	82.03	74.18	77.90
SpanishBERT	80.9	72.02	76.20
CamemBERT	81.91	71.45	76.32
GermanBERT	80.61	71.43	75.74
JapaneseBERT	79.56	65.36	71.76
HindiBERT	79.56	65.68	71.95
ThaiBERT	79.11	66.25	72.11
IndonesianBERT	78.96	65.66	71.69
CMTA	81.52	74.40	77.79

Table 7: Language-wise predicted misinformation labels of tweets in February, March and April.

Lingo	**February**			**March**			**April**		
	Misinformation			**Misinformation**			**Misinformation**		
	False	**Partially False**	**Misleading**	**False**	**Partially False**	**Misleading**	**False**	**Partially False**	**Misleading**
Spanish	58346	6653	13740	67956	10913	8826	34125	5437	3604
German	517	581	2505	862	1438	3043	584	892	2664
Japanese	1920	3079	5245	448	692	2650	1635	2850	5840
Indonesian	11157	3226	1951	12573	4336	1582	9073	3367	1273
English	88369	62747	76640	92428	96571	105143	77368	74947	63473
French	4464	3472	1155	12024	10270	1670	6650	5300	763
Hindi	500	870	202	756	909	348	2211	2868	705
Thai	1950	1074	2780	6036	736	7678	2263	554	2917

It showcases the overall (all 5 months together) spread of misinformation types across each language. We could see that German tweets have the highest number of 'Misleading' tweets, whereas French have the least. Spanish tweets beat other language's tweets by becoming the largest source of 'False' misinformation. Germany generated the least number of 'False' tweets. Hindi tweets tend to have the highest number of 'Partially false' tweets, whereas Thai have the least.

We could observe that for February, March and June months, our model predicted a large number of tweets as 'False', followed by 'Misleading', which is the second largest and the number of 'Partially false' was the least (see Figure 3). Our model discovered that the number of 'Partially false' tweets are more than 'Misleading' tweets and 'False' tweets were again in the majority for the tweets generated during April and May.

Table 7 and 8 present a detailed count of misinformation classes across all the languages. The following specific observations were made concerning the languages: The misinformation distribution for English data indicates that there is a majority of **False** tweets during the five months, whereas the distribution of **Misleading** labelled data is slightly less than as compared to **False** labelled data. **Partially False** labelled tweets are moderately distributed, as in the month April, we can see that there is a more significant number concerning other months. According to the language wise-distribution shown in Figure 2, Spanish tweets have a greater frequency of **False** labelled tweets, whereas the



Fig. 2: Month-wise Disinformation Distribution in Languages.

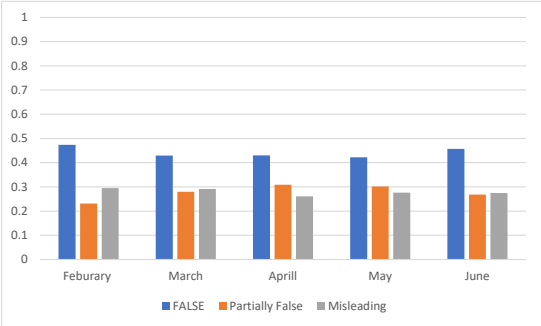


Fig. 3: Month-wise Disinformation Distribution.

Table 8: Language-wise predicted misinformation labels of tweets in May and June.

Lingo	May			June		
	Misinformation			Misinformation		
	False	Partially False	Misleading	False	Partially False	Misleading
Spanish	57821	8214	7107	54965	8828	6759
German	1076	1426	4430	616	657	2028
Japanese	8984	12324	18125	1741	2496	3389
Indonesian	12695	4574	1805	9114	3038	1000
English	140494	128326	119391	135172	101896	109483
French	8475	7667	842	4952	3535	483
Hindi	4560	6057	1343	2501	2739	751
Thai	2825	470	1830	2103	486	3122

Misleading tweets and **Partially False** tweets shows the almost identical number of tweet across the five months. There was a surge of **Misleading** labelled tweets during February, and the count remained the same throughout the five months. There was also an increase in **Partially False** tweets in March, but it decreased in successive months, leading to minor **False** labelled tweets. According to the language wise-distribution shown in Figure 2, on average throughout the five months, approx 20% of Japanese tweets are labelled **False**. Similarly, approx 30% of the Japanese tweets are labelled **Partially False**, leading to the majority of 50% data are labelled as **Misleading**. We can also see a considerable increase in **Misleading** tweets in March, tweeted in the Japanese language. According to the distribution of Indonesian tweets, approximately 10% of tweets are labelled as **Misleading**, and on the contrary, there is a large distribution of **False** labelled tweets. Approximately 34% of the Indonesian dialect data is labelled as **Partially False** throughout the five months. Figure2 shows the misinformation distribution across all five months in the French tweets. The largest majority of the tweets were classified as **False** misinformation. Among **Partially false** and **Misleading**, the least number of tweets were labelled as **Misleading**. The frequency of Hindi tweets is low in the dataset used in our experiment. However, our model can predict or label Hindi tweets. Tweets in

Hindi have low numbers of **Misleading** tweets, whereas the **Partially False** tweets class has a great frequency. **False** labelled tweets are slightly low compared to **Partially False** tweets in this dialect. The distribution of Thai tweets shows that our model prediction is majorly oriented towards the **Misleading** tweets. The distribution of **Misleading** labelled tweets is the greatest among the labelled classes, in contrast to **Partially False** tweets. **False** labelled tweets are comparatively moderate in this language.

5 Conclusion and Future Work

This paper introduced CMTA, a multilingual model for analyzing text applied to classify COVID-19 related multilingual tweets into misinformation categories. We demonstrated that our multilingual CMTA framework performed significantly well compared to the monolingual misinformation detection models used independently. Experimental validation of CMTA detected misinformation distribution across eight significant languages. The paper presented a quantified magnitude of misinformation distributed across different languages in the last five months. Our future work aims to collect more annotated training data and perform analysis of a larger multilingual dataset to gain a deeper understanding of misinformation spread. We are currently improving our model's robustness and contextual understanding for better performance in the classification task. We hope that researchers could gain deeper insights about misinformation spread across major languages and use the information to build more reliable social media platforms through our work.

References

1. Alam, F., Shaar, S., Dalvi, F., Sajjad, H., Nikolov, A., Mubarak, H., Martino, G.D.S., Abdelali, A., Durrani, N., Darwish, K., Nakov, P.: Fighting the covid-19 infodemic: Modeling the perspective of journalists, fact-checkers, social media platforms, policy makers, and the society (2020)
2. Brennen, J.S., Simon, F., Howard, P.N., Nielsen, R.K.: Types, sources, and claims of covid-19 misinformation. Reuters Institute **7** (2020)
3. Brindha, M.D., Jayaseelan, R., Kadeswara, S.: Social media reigned by information or misinformation about covid-19: a phenomenological study (2020)
4. Chen, E., Lerman, K., Ferrara, E.: Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus twitter data set. JMIR Public Health and Surveillance **6**(2), e19273 (2020)
5. Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C.M., Brugnoli, E., Schmidt, A.L., Zola, P., Zollo, F., Scala, A.: The covid-19 social media infodemic. arXiv preprint arXiv:2003.05004 (2020)
6. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
7. Eysenbach, G., Powell, J., Kuss, O., Sa, E.R.: Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. Jama **287**(20), 2691–2700 (2002)

8. Frenkel, S., Alba, D., Zhong, R.: Surge of virus misinformation stumps facebook and twitter. The New York Times (2020)
9. Gencoglu, O., Gruber, M.: Causal modeling of twitter activity during covid-19. arXiv preprint arXiv:2005.07952 (2020)
10. Hernon, P.: Disinformation and misinformation through the internet: Findings of an exploratory study. Government information quarterly **12**(2), 133–139 (1995)
11. Huang, B., Carley, K.M.: Disinformation and misinformation on twitter during the novel coronavirus outbreak. arXiv preprint arXiv:2006.04278 (2020)
12. Ingraham, N.E., Tignanelli, C.J.: Fact versus science fiction: fighting coronavirus disease 2019 requires the wisdom to know the difference. Critical Care Explorations **2**(4) (2020)
13. Kouzy, R., Abi Jaoude, J., Kraitem, A., El Alam, M.B., Karam, B., Adib, E., Zarka, J., Traboulsi, C., Akl, E.W., Baddour, K.: Coronavirus goes viral: quantifying the covid-19 misinformation epidemic on twitter. Cureus **12**(3) (2020)
14. Lopez, C.E., Vasu, M., Gallemore, C.: Understanding the perception of covid-19 policies by mining a multilanguage twitter dataset. arXiv preprint arXiv:2003.10359 (2020)
15. Organization, W.H., et al.: Coronavirus disease 2019 (covid-19): situation report, 188 (2020)
16. Ortiz-Ospina, E.: The rise of social media (2020), https://ourworldindata.org/rise-of-social-media
17. Pennycook, G., McPhetres, J., Zhang, Y., Lu, J.G., Rand, D.G.: Fighting covid-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. Psychological science **31**(7), 770–780 (2020)
18. Poynter Institute, .: The international fact-checking network. (2020), https://www.poynter.org/ifcn/
19. Saire, J.E.C., Navarro, R.C.: What is the people posting about symptoms related to coronavirus in bogota, colombia? arXiv preprint arXiv:2003.11159 (2020)
20. Shahi, G.K., Nandini, D.: Fakecovid–a multilingual cross-domain fact check news dataset for covid-19. arXiv preprint arXiv:2006.11343 (2020)
21. Sharma, K., Seo, S., Meng, C., Rambhatla, S., Liu, Y.: Covid-19 on social media: Analyzing misinformation in twitter conversations. arXiv preprint arXiv:2003.12309 (2020)
22. Singh, L., Bansal, S., Bode, L., Budak, C., Chi, G., Kawintiranon, K., Padden, C., Vanarsdall, R., Vraga, E., Wang, Y.: A first look at covid-19 information and misinformation sharing on twitter. arXiv preprint arXiv:2003.13907 (2020)
23. Wannaphong Phatthiyaphaibun, Korakot Chaovavanich, C.P.A.S.L.L.P.C.: PyThaiNLP: Thai Natural Language Processing in Python (Jun 2016). https://doi.org/10.5281/zenodo.3519354, http://doi.org/10.5281/zenodo.3519354
24. Wilford, J., Osann, K., Wenzel, L.: Social media use among parents of young childhood cancer survivors. Journal of Oncology Navigation & Survivorship **9**(1) (2018)
25. Wu, Y., Schuster, M., Chen, Z., Le, Q.V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., et al.: Google's neural machine translation system: Bridging the gap between human and machine translation. arXiv preprint arXiv:1609.08144 (2016)