

Data and models for stance and premise detection in COVID-19 tweets: insights from the Social Media Mining for Health (SMM4H) 2022 shared task

Vera Davydova^a, Huabin Yang^b, Elena Tutubalina^{c,d,a}

^a*Sber AI, 19 Vavilova St., Moscow, Russian Federation, 117997*

^b*University of Electronic Science and Technology of China, Qingshuihe Campus:No.2006, Xiyuan Ave, West Hi-Tech Zone, Chengdu, China, 611731*

^c*Kazan Federal University, 18 Kremlyovskaya street, Kazan, Russian Federation, 420008*

^d*National Research University Higher School of Economics, 11 Pokrovsky Bulvar, Moscow, Russian Federation, 109028*

Abstract

The COVID-19 pandemic has sparked numerous discussions on social media platforms, with users sharing their views on topics such as mask-wearing and vaccination. To facilitate the evaluation of neural models for stance detection and premise classification, we organized the Social Media Mining for Health (SMM4H) 2022 Shared Task 2. This competition utilized manually annotated posts on three COVID-19-related topics: school closures, stay-at-home orders, and wearing masks. In this paper, we extend the previous work and present newly collected data on vaccination from Twitter to assess the performance of models on a different topic. To enhance the accuracy and effectiveness of our evaluation, we employed various strategies to aggregate tweet texts with claims, including models with feature-level (early) fusion and dual-view architectures from the SMM4H 2022 Task 2 leaderboard. Our primary objective was to create a valuable dataset and perform an extensive experimental evaluation to support future research in argument mining in the health domain.

Keywords: natural language processing, COVID-19 tweets, opinion mining, argument mining, stance detection, premise detection

Email addresses: veranchos@gmail.com (Vera Davydova), 202121080420@std.uestc.edu.cn (Huabin Yang), ElVTutubalina@kpfu.ru (Elena Tutubalina)

1. Introduction

In recent years, social media platforms have become powerful channels for public discourse, shaping opinions, and disseminating information on various topics. The advent of the COVID-19 pandemic has further intensified the prominence of online discussions, particularly regarding critical public health measures such as masks, lockdowns, school closures, and vaccine mandates. Understanding the stances and premises expressed in these discussions is crucial for policymakers, healthcare professionals, and researchers to gauge public sentiment, identify misinformation, and develop effective communication strategies [1, 2].

This paper addresses the challenge of automatic stance and premise detection in tweets specifically related to COVID-19 mandates. Stance detection involves determining the point of view (stance) of the text’s author towards a particular topic, while premise detection aims to identify the underlying reasons for supporting those stances. By automatically analyzing a large volume of tweets, we can gain valuable insights into the prevailing opinions, concerns, and rationales within the online community. Detecting stances and premises in tweets poses unique challenges due to the limited length of the messages, the informal nature of the language used, and the presence of noise and ambiguity. Furthermore, the topic of COVID-19 mandates is highly polarized, with divergent perspectives ranging from enthusiastic support to vehement opposition. Therefore, developing robust and accurate computational models capable of capturing the nuanced stance and premises expressed in tweets is vital for gaining a comprehensive understanding of the public discourse surrounding this critical issue.

A preliminary version of this work has appeared in [3, 4]. In this journal version, we have made several significant improvements, including:

1. Annotation of an external test dataset on a new claim topic related to vaccination. This topic was not included in the training set of the SMM4H

2022 Task 2, which covered three other claims: school closures, stay-at-home orders, and wearing masks. The dataset is annotated by human experts, ensuring high-quality labels for training and evaluation purposes.

2. An extended description of the experimental datasets and emotion analysis of tweets.
3. An extended description of the high-scoring systems used in the SMM4H 2022 Task 2. These models combine deep learning algorithms and linguistic features to predict stances and identify premises within the tweet.
4. Investigation of model performance on different claim topics, with the addition of new experimental results and conclusions.
5. Error analysis of the best-performing model and a discussion of its limitations.

In this work, we seek to answer the following research questions: **RQ1:** To what extent can models trained on specific claims be transferred to generalize and apply to other claims within the same domain? **RQ2:** Can the fusion of tweets and corresponding claims significantly enhance the performance of models?

The remainder of this paper is organized as follows. Sec. 2 provides an overview of related work on stance and premise detection in social media. Sec. 3 describes the methodology employed in the data collection and annotation process. In Sec. 4.4, we present the experimental setup, model architectures, and evaluation metrics. Sec. 5 describes obtained results. Finally, in Sec. 6, we discuss the implications of our findings, limitations of the study, problems that we faced during data collection, and potential directions for future research. All the data and code written in support of this publication are publicly available via https://github.com/Veranchos/ArgMining_tweets.

2. Related Work

2.1. Argument Mining on COVID stances

In recent years, there has been a growing interest in understanding public sentiment and stance on social media platforms, particularly regarding public health issues such as COVID-19. Several studies have focused on analyzing tweets to classify the stances expressed by users toward various aspects of the pandemic. This section provides an overview of the key findings and methodologies employed in previous research related to stance classification in tweets about COVID-19.

Since the beginning of the COVID-19 pandemic, new datasets and models in health-related issues have been created [5, 6, 7, 8]. [6] presented a large-scale curated dataset of over 1.12 billion tweets, growing daily, related to COVID-19 chatter. The largest dataset of Twitter users’ stances in the context of the COVID-19 pandemic is COVID-CQ [9]. It consists of controversial tweets about the efficacy of hydroxychloroquine as a treatment. Similarly, [7] presented a dataset for biomedical claim detection in Twitter posts. [8] created the LockdownTweets – mostly unlabelled tweet dataset, related to the lockdown policy in New York State during the pandemic. People’s opinions towards health mandates in Germany are investigated in [10]: first, relevant tweets were identified, and then the expressed stances were detected. While most researchers concentrate on the stance detection task, a few datasets exist for premise classification. [11] gathered a collection of short texts from Russian social media, namely posts’ commentaries. These data contain both statements defining the author’s stance towards the given claims, and statements with premises “for” / “against” these claims. [12] also explored the premises and argumentation behind the anti-masks tweets. They analyzed 267k tweets about mask-wearing and identified the most common reasons for opposition to masks mandates using machine learning and qualitative content analysis. Even before the first vaccines for COVID-19 treatment were discovered, and especially with the release of vaccination programs all over the world, obligatory vaccine mandates became the central topic in

social media. A few datasets in different languages with people’s reactions to vaccination orders exist at the moment: [13] for Indonesian/Bahasa, [14] for Italian, Arabic [15]. Most of the existing works in this area are dedicated to detecting the spreading of misinformation about vaccination. It is worth noting that while existing research has made significant contributions to stance classification in the context of COVID-19 tweets, challenges and limitations persist. Some studies have acknowledged the issue of ambiguous or conflicting stances expressed within tweets, highlighting the need for more nuanced approaches to capture the complexity of user opinions accurately.

3. Dataset

The dataset contains 8236 tweets and their stances and premises related to COVID-19 mandates. 6166 tweets from this data set are designated to three claims: support or opposition to face masks, school closures, and stay-at-home orders. This collection was leveraged in our preliminary work for the SMM4H 2022 Task 2 [3]. The remaining 2070 tweets are attributed to vaccine mandates and were collected and manually annotated within this work.

The training part of the data set is based on [16] and consists of 3,556 tweets. There is a balanced mix of three topics: 37%, 33%, and 30% of the tweets are about face masks, school closures, and stay-at-home orders, respectively. Both validation and test sets were collected using Twitter API. Then, the preprocessing algorithms were applied to remove tweets that were too short, contained only hashtags or mentions, or were likely considered advertisements or spam. The data collection pipeline is illustrated on Fig. 1.

3.1. Vaccination tweets

We also wanted to check whether we can leverage models trained on the data from [3] to predict stances and premises related to a slightly different topic without additional training. Thus, we included tweets about an additional COVID-related mandate in the test set: vaccination. To collect the relevant tweets about COVID-19 vaccine mandates, we proceeded with the same

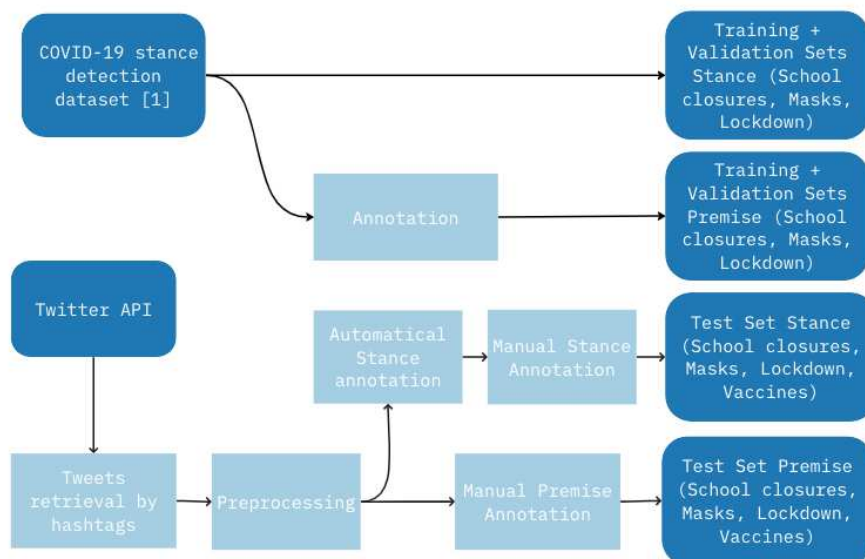


Figure 1: Data collection pipeline.

approach as in our previous work related to other COVID-19 mandates. Firstly, we analyzed existing works and datasets regarding tweets about vaccine attitudes. Then, we extracted relevant hashtags and collected tweets written between January 2021 and November 2022 containing these hashtags. Since our work is dedicated to the detection of people’s stances towards COVID-related mandates, we additionally analyzed all the extracted keywords used in those tweets and allocated keywords regarding people’s stances towards vaccine mandate (i.e., `#NoMandatoryVaccines` or `#vaccinatetheworld`). We also manually annotated the stances of those hashtags. The complete list of hashtags and their stances can be found in Appendix .7. Although hashtags are great for selecting likely relevant tweets; they are noisy and not reliable enough to accurately identify the stance towards a target for a tweet. Since textual data collected from social media suffers from inevitable noises, such as emoticons, URLs, and misspellings, pre-processing is performed to reduce the impact of noise. Specifically, the URLs, mentions, and hashtags were removed by using the python

package *tweet-preprocessor*¹, and the emoticons are converted into their textual representations by using the python package *emoji*². We also removed duplicates and short tweets (less than 150 symbols). Finally, we randomly selected 3000 tweets with an equal proportion of tweets related to different automatically labeled stances (favour/against/neither). The resulting tweets were used for annotation.

3.2. Collection and Annotation

We followed annotation guidelines of an argument mining shared task RuArg-2022 [11]. Below, we highlight some of the key features of our guidelines:

- A statement is evaluated as an argument if it contains a statement that can be used in a dispute to persuade an opponent. For example, *masks help prevent the spread of the disease.* (1)
- It is also necessary to distinguish sentiment (positive and/or negative) from argumentation. For example, *and the fact that Trump did not introduce a suffocating quarantine is well done!* (0)
- The argument should not be a fragment that needs to be thought out. For example, *It is effective if you declare a quarantine.* (0)
- An example of an argument could be such a common sense statement. For example, *in all countries of the world, everyone is wearing masks, but ours... this is not a joke.* (1)
- The position of the author “favor” or “against” should be clear – only under this condition it is possible to detect an argument. The annotator should not think for the author. For example, the author’s position on quarantine is unclear in the text *here are the words of my classmate*

¹<https://pypi.org/project/tweet-preprocessor/>

²<https://pypi.org/project/emoji/>

from Annecy, France, from today’s Facebook correspondence - “France introduced quarantine, and immediately everyone poured out to barbecue in nature.” (0).

Annotators had two tasks: 1) Check whether the automatically labeled tweet’s stance is correct. If it was incorrect, they needed to change it to the correct one. 2) Annotate whether the tweet contains a premise according to stance: label 1 means the tweet contains a premise, and label 0 means the absence of a premise. All tweets were annotated by five annotators, then we selected only those tweets that had a high degree of agreement between annotators (minimum 4/5 annotations were identical for each subtask). As a result, we created a new dataset containing 2070 tweets about vaccination, which was later used to test the models. As a crowd-source platform, we use *Yandex.Toloka*³ for the SMM4H 2022 Task 2 and *TagMe*⁴ for tweets about vaccination. Tab. 1 illustrates the examples from the resulting data set.

3.3. Analysis of the annotated dataset

Tab. 2 shows statistics of experimental datasets across topics. The training set includes 38% and 25% in-favor and against tweets, respectively. Relative class balance is also present for argumentation: 63% of train tweets contain a premise, 34% of tweets in the test set contain a premise; 26% of tweets in the test set are annotated as in-favor. As shown in Tab. 2, the distribution of classes by topic is different. Thus, the topic of staying-at-home orders contains more “against” tweets than tweets “in-favor”, and a neutral stance is much more common for tweets about vaccination. In almost every claim, there are fewer premises in tweets.

3.4. Emotion Analysis

In addition to the stance and premise annotation, our research aimed to investigate the emotions expressed in tweets concerning vaccination and determine

³<https://toloka.yandex.ru/>

⁴<https://tagme.sberdevices.ru/>

Tweet	Stance	Premise
"The first and most powerful, tool we have against COVID-19 is vaccination. Vaccination is by far the most important way to protect ourselves against severe illness, hospitalisation or death in the event that you contract the coronavirus #IChooseVaccination #VaccinesSaveLives "	favor	1
100% this man.. Vaccine passports have failed dismally in all countries who tried it. They only serve to provide a false sense of security, discrimination and segregation. #prochoice #NoVaccinePassportAnywhere	against	1
They're trying to do that with the fake vaccine anyway. A MANDATED vaccine for a 99.9% survival rate is nuts. They've managed to convince the whole world they're going to die if they're not coerced into it. That's fishy. Agenda 2030 is in effect.	against	1
Biden's vaccine mandates are the most egregious U.S. human rights violations since slavery. Biden is another Josef Mengele, the Nazi "Angel of Death." We must fight against these terrible injustices, and vote out all Democrat Party politicians. #novaccinemandates #impeachBiden	against	0
It's like soon as the tv comes on the 1st thing I hear is either covid, vaccine or booster it's like they have nothing else to talk about, sure enough it's turned off within minutes	neither	0
People infected with the omicron variant were almost 60 percent less likely to enter the hospital than those infected with delta. Regardless odds are better with vax+ boost. Government should be more active in advertising vaccination.	favor	1
Santa popped by one of our south west London vaccination centres to say thank you to all the wonderful vaccinators and volunteers! Get your Covid vaccination for Christmas!	favor	0

Table 1: Examples of tweets about COVID-19 vaccine mandates annotated for stance and premise classification.

whether specific emotions correlate with the authors' positions on mandatory vaccination. We employed the emotion classification tool developed by [17] to address this inquiry.

Claim/Topic	Stance			Premise	
	favor	against	neither	1	0
the SMM4H 2022 Task 2: train set (3566 tweets)					
face masks	652	324	343	508	811
close school	526	217	307	535	515
home orders	168	333	686	288	899
the SMM4H 2022 Task 2: validation set (600 tweets)					
face masks	121	51	36	82	126
close school	91	35	51	80	97
home orders	32	72	111	58	157
the SMM4H 2022 Task 2: test set (2000 tweets)					
face masks	209	208	260	253	424
close school	215	192	263	294	376
home orders	102	170	381	169	484
new annotated tweets (2070 tweets)					
vaccines	421	279	1370	614	1456

Table 2: Summary of statistics of stance and premise classification datasets. The topic on *school closures* and *stay at home orders* has been shortened to *close school* and *home orders*, respectively.

The employed model, DistilRoBERTa-base, was fine-tuned on a balanced subset derived from six diverse data collections. This model predicts Ekman’s six basic emotions [18], along with a neutral class. Fig. 2 illustrates the distribution of emotions observed in the collected tweets. Tab. .6 in the appendix illustrates some examples of tweets labeled according to the emotions expressed.

Anger emerged as the predominant emotion among all combined tweets. This observation suggests that the topic of vaccination elicits intense emotions among users. Furthermore, this finding aligns with previous research indicating that anger is a prevalent emotion in tweets pertaining to COVID-19 [19]. Moreover, studies have shown that tweets provoking anger are more likely to be

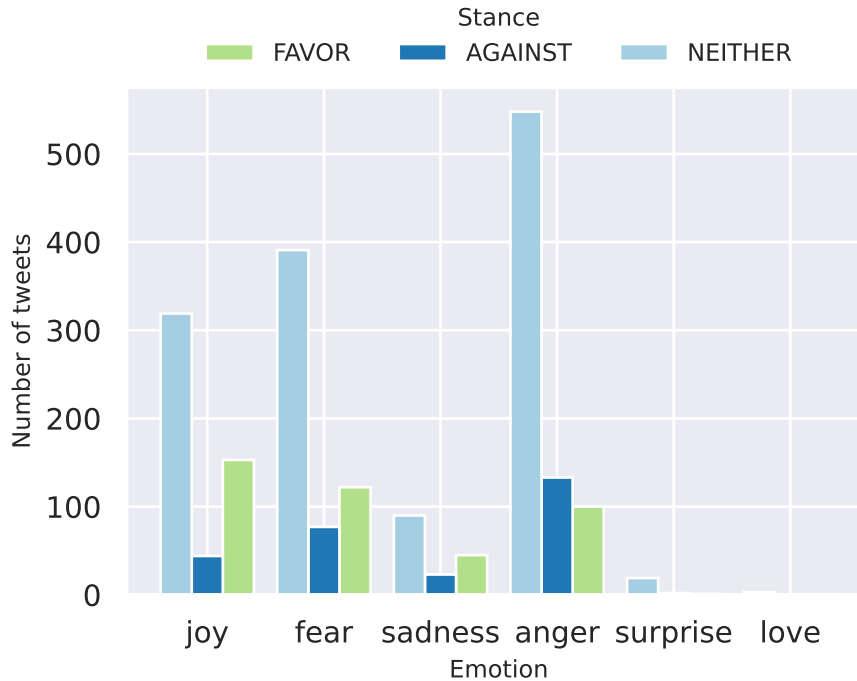


Figure 2: Distribution of 6 most common emotions in tweets per stance.

shared [20, 21]. Notably, anger is predominantly expressed in tweets authored by individuals who oppose mandatory vaccination. During the annotation process, we observed that tweets adopting an opposing stance often exhibited aggressive rhetoric and insults directed towards supporters of the opposing viewpoint. While such tendencies were also observed in tweets expressing support for mandatory vaccination, they were relatively less frequent.

Another notable emotion observed in tweets with the “FAVOR” label was joy. Users frequently expressed joy regarding vaccine development, shared successful vaccination experiences, and expressed hope for the prompt resolution of the pandemic.

Furthermore, fear was found to be one of the prevalent emotions in tweets about vaccination. It is essential to mention that during our analysis, we noted instances where the model incorrectly assigned the label “fear” to more neutral tweets. We attribute this misclassification to the inherent association between

the broader topic of a pandemic and the emotion of fear, potentially leading to a slight bias in the model’s predictions. To mitigate this issue, we plan to incorporate manual labeling and conduct a more comprehensive analysis using data labeled by domain specialists in future research.

In summary, the most common emotions observed in tweets about mandatory vaccination are anger, fear, and joy. This finding underscores the relevance and high emotional sensitivity surrounding this topic. A deeper analysis of these emotions can contribute to the development of stance detection systems, and the insights gained can inform public opinion monitoring systems.

4. Models

Most teams utilized COVID-related BERT models for the SMM4H 2022 Task 2, along with additional techniques like regularized dropout (R-drop), to address issues with unbalanced label distribution and overfitting; see [4] for F1 scores for each of the 14 team’s best-performing system for both tasks.

In this paper, we focus on three key model characteristics: (i) aggregation of claim and tweet texts as input data of a language model, (ii) syntactic features, and (iii) dual-view architecture that learns the representations of both types of texts. In this regard, we employed models that proved their robustness for stance and premise detection in tweets related to COVID-19 and were ranked as the top-5 best-performing architectures during the SMM4H 2022 Task 2: BART + syntax features [22] and DAN-Bert [23]. Each model leveraged different techniques to capture the nuanced information in the tweets. All the used models were trained in two set-ups: basic (utilizing the tweets only) and with the adoption of the early fusion technique [24]. Early fusion employs concatenated features from two modalities (tweets texts and corresponding claims) as input features for the classifiers. The models used in our research are described below.

4.1. BERT-base model

Since Transformer-based models [25] produce robust context-based representations of textual data, we leverage BERT-base-uncased as our base model to

solve the premise and stance detection task. This is a BERT model consisting of 12 layers of Transformer encoder, 12 attention heads, 768 hidden size, and 110M parameters [26]; we fine-tuned the model for the text classification tasks on our data.

4.2. BART + Syntax features

One of the best-performing models in the SMM4H 2022 Task 2 leveraged the syntactic features of the text. Thus, we included this model in the comparison. This model is based on BART architecture [27], an encoder-decoder transformer model with a BERT-like bidirectional encoder and auto-regressive decoder like GPT. Since BART’s pre-training task encourages the model to learn representations that are robust to noise and variations in the input text, the model is well-suited for tasks that require handling noisy and ambiguous text. To capture the syntactic structure of the input sentences, we incorporate a feature obtained from the dependency parse tree of the sentence. We leverage the open-source tool Spacy [28] to construct the dependency parse tree. Following that, the dependency features are encoded in descending order of their occurrence. Consider three distinct dependency tags: ‘aux’, ‘amod’, ‘nsubj’ in all of the input sentences in the original dataset. If ‘aux’, ‘nsubj’, and ‘amod’ have their occurrence count in the order mentioned, then ‘aux’, ‘nsubj’, and ‘amod’ are encoded as one, two, and three, respectively. These encoded features are concatenated to the transformer model’s output and fed into the classifier head to obtain the stance/premise. Instead of the usual cross-entropy loss, the supervised contrastive loss function [29] was used. This helped leverage label information better and morph the embedding space by pulling data points from the same class closer and pushing apart data points from other classes. The dataset for contrastive training was created by adding positive and negative labels to the samples present in the task dataset. We try three different strategies: fine-tuning with contrastive loss only, pre-training with contrastive loss, then fine-tuning with cross-entropy loss, and fine-tuning with a weighted loss with 0.7 weight given to cross-entropy and 0.3 to contrastive loss. The weights

were chosen considering the performance when fine-tuning with cross-entropy and contrastive loss.

4.3. COVID-Twitter-BERT

BERT-large-uncased model, pre-trained on 97M unique tweets (1.2B training examples) collected between January 12 and July 5, 2020, containing at least one of the keywords “wuhan”, “ncov”, “coronavirus”, “covid”, or “sars-cov-2”. These tweets were filtered and preprocessed to reach a final sample of 22.5M tweets (containing 40.7M sentences and 633M tokens), which were used for training [30]. We fine-tuned this model for our two classification tasks: binary classification for Premise detection and 3-class classification for Stance prediction. The following parameters were leveraged during this fine-tuning. Cross-entropy was leveraged as a loss function with the learning rate set to 4,00E-05. We also used AdamW [31] as an optimizer with the weight decay set to 0. We trained the model for 10 epochs with the batch size set to 8 and the maximum sequence length set to 128.

4.4. DAN-BERT

This architecture draws inspiration from the dual-view adaptation neural network [32], which learns both representations of subjective and objective features of texts. Fig. 3 presents the overall system architecture.

Given a tokenized tweet, the output embedding of [CLS] token is obtained through the COVID-Twitter-BERT-v2 model [30], which is a pre-trained BERT model on a large corpus of tweets regarding COVID-19 and is the latest version with better downstream performance ⁵. [CLS] is a special classification token and the last hidden state of BERT which is used for classification tasks. We employ it because of its ability to capture a representation of the whole sequence.

For extracting subjective and objective information separately, the [CLS] token is duplicated and passed through two COVID-Twitter-BERT-v2 models

⁵<https://huggingface.co/digitalepidemiologylab/covid-twitter-bert-v2>

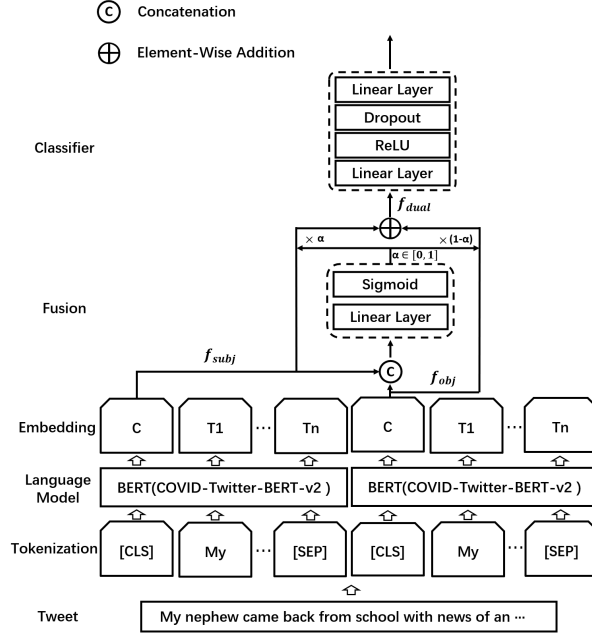


Figure 3: The overall dual-view architecture. The text is tokenized with a [CLS] token and a [SEP] token added at the beginning and end, respectively. The tokens are passed through the pre-trained language models to get their corresponding embeddings, followed by a fusion block and a classifier.

to obtain the corresponding embeddings (i.e., subjective and objective feature vectors, denoted by f_{subj} and f_{obj} , respectively).

These two embeddings are then concatenated and fed into a fusion block composed of a fully connected layer with hidden dimensions of 2048 and a sigmoid activation function to produce a weight vector as follows,

$$\alpha = \text{sigmoid}(\mathbf{W}[f_{subj}, f_{obj}] + \mathbf{b}), \quad (1)$$

where $[,]$ denotes vector concatenation and \mathbf{W}, \mathbf{b} are trainable parameters.

The goal of the fusion block is to attain an optimal combination between f_{subj} and f_{obj} . Once the weight vector α is obtained, the weighted sum of f_{subj}

and \mathbf{f}_{obj} can be computed as follows,

$$\mathbf{f}_{dual} = \alpha \odot \mathbf{f}_{subj} + (1 - \alpha) \odot \mathbf{f}_{obj}, \quad (2)$$

where \odot denotes the element-wise product and \mathbf{f}_{dual} represents the weighted sum of \mathbf{f}_{subj} and \mathbf{f}_{obj} .

Finally, the vector \mathbf{f}_{dual} is passed through a classifier, which consists of two linear layers with hidden dimensions of 1024, a ReLU activation function, and a dropout layer with a dropout probability of 0.15.

Implementation. In an implementation, there are 50 training epochs, the batch size is set to 16, the optimizer is AdamW [33] with a learning rate of 1e-6 and a weight decay rate of 0.01, the max length of input is set to 128, and the loss function is cross-entropy loss.

Additionally, cross-validation is a data resampling method used to evaluate models and prevent overfitting. The basic form of cross-validation is k-fold cross-validation [34]. Therefore, the 5-fold cross-validation with a majority voting strategy is applied in the implementation and also proved to improve the performance of the proposed system.

All the experiments are conducted on a server with an Inter Xeon Gold 6142M CPU and an NVIDIA GeForce RTX 3090 GPU.

5. Results

5.1. Evaluation

We leveraged F_1 as the main evaluation metric in each of the two subtasks, which is calculated according to the following formula: $F_1 = \frac{1}{n} \sum_{c \in C} F_{1_{rel_c}}$, where $C = \{\text{“face masks”}, \text{“stay at home orders”}, \text{“school closures”}\}$, n is the size of C , $F_{1_{rel_c}}$ is macro F_1 -score averaged over two classes for each task (in-favor & against classes for stance; 0 & 1 classes for premise). The F_1 is also used for evaluation results obtained on vaccine mandates data set. Systems performance for data related to school closures, masks, and stay-at-home orders (topics that were leveraged during training processes) is described in Tab. 3.

Separately, we evaluated the best-performing models on the data related to vaccine mandates (Tab. 4).

All the models perform significantly better than the random baseline. COVID-Twitter Bert shows the best results on the test set containing tweets about school closures, stay-at-home orders, and masks mandates. This is a large-scale model, trained specifically on the tweets about COVID-19; therefore, it shows the best results on such texts. However, this model performed second best on vaccination data. We attribute this to the fact that COVID-Twitter BERT was trained on a corpus of 160M tweets about the coronavirus collected during the period from January 12 to July 5, 2020. This time dates back to the beginning of the pandemic when vaccines had not yet been developed and corresponding mandates had not been announced. Complex models leveraging syntax features (BART+syntax features) or dual architectures (DANBert) work much better than a general-domain language model (BERT). Their performance on the SMM4H data set is close to that of models trained on specific Twitter data (COVID-Twitter BERT). The best results on the new data set containing data about mandatory vaccination are demonstrated by DANBert.

The difference between the models' performances is statistically significant (Wilcoxon p -value ≤ 0.05) [35] on the two datasets. However, models trained on the the SMM4H 2022 Task 2 data significantly outperform a random baseline on the vaccination dataset. It indicates that they are capable of providing useful insights and predictions on the unseen topic and can be successfully transferred to the tweets in the same domain but in slightly different theme. Moreover, adding tweet claims to the text during training can improve the performance on the unseen data in the new domain (vaccines). Nevertheless, it does not affect the results on the data in the same domains it was trained.

Fig. .4 presents the bar plots demonstrating the F1 metric for each model per each claim.

Model	Tweets		Tweets + Claims	
	F ₁ Stance	F ₁ Premise	F ₁ Stance	F ₁ Premise
Random	0.268	0.33	0.268	0.33
General-domain BERT	0.464	0.352	0.446	0.354
COVID-Twitter-BERT	0.601	0.719	0.245	0.68
BART+Syntax	0.45	0.34	0.38	0.35
DANBert	0.581	0.704	0.581	0.684

Table 3: F₁-scores (F₁) for stance and premise detection in tweets about school closures, masks, and lockdown

Model	Tweets		Tweets + Claims	
	F ₁ Stance	F ₁ Premise	F ₁ Stance	F ₁ Premise
Random	0.223	0.286	0.223	0.286
General-domain BERT	0.26	0.568	0.298	0.57
COVID-Twitter-BERT	0.255	0.555	0.36	0.456
DANBert	0.41	0.592	0.43	0.61

Table 4: F₁-scores (F₁) for stance and premise detection in tweets about vaccines.

6. Discussion

During the process of dataset creation, certain data characteristics emerged that posed challenges and led to disagreements among annotators. These instances also proved to be perplexing for language models (LMs). It is essential to acknowledge that these cases may hinder the accurate assessment of public opinion regarding mandatory vaccination. In this section, we will provide a brief description of these challenging cases. Firstly, we encountered tweets that strongly endorsed vaccines; however, the author’s stance on mandatory vaccination remained ambiguous. Tab. 5 presents examples illustrating such cases. Additionally, a considerable number of tweets consisted of news reports or statistical information. Despite implementing algorithms to filter out such texts, they still appeared in the datasets since they were not solely posted by news

Tweet text	Problem
<i>Lots of cat content. Despite the vaccine, I am feeling the worst I have yet with this Breakthrough Covid case. Isolated in my room from my husband, sons, fam for Christmas, at least I have my Edna</i>	unclear position
<i>You really are an insensitive imbecile grow up and respect the fact that real people are hurting here leave your anti vax posts to those that care</i>	emotion
<i>Day 645: #COVID19 cases keep rising! The DeltaVariant and OmicronVariant cases keeps rising! WearAMask GetVaccinated GetBoosterShot global: Cases: 277,167,932 Deaths: 5,377,435 US: Cases: 51,545,991 Deaths: 812,069</i>	statistics
<i>How many politicians do you think have these vaccine companies in their back pockets? You don't become a millionaire by joining Congress without backdoor deals.</i>	conspiracy theory
<i>Dear sweet, innocent Marie. It's time to break you of your bubble. I'm sorry I'm the one that has to be it. #COVIDIOTS #antivaxxers #coronavirus #Unvaccinated #GetVaccinated</i>	reply

Table 5: Examples of tweets containing ambiguous stance.

accounts but also by real users. Moreover, certain tweets implicitly conveyed the author’s position, requiring additional contextual information for accurate stance assessment. For instance, the last row in Tab. 5: a tweet reacting to another user’s remark without access to the original post can challenge determining the stance expressed. Furthermore, emotional tweets were prevalent, which could be perplexing as the emotion was evident, but the stance was not explicitly presented. The provided example (second row in Tab. 5) showcases clear aggression, allowing us to infer the author’s support for vaccination. However, the tweet does not address the attitude towards mandatory vaccination. Additionally, a substantial number of tweets expressed support for various conspiracy theories about vaccination. However, even in these cases, the author’s stance on COVID-19 mandates is not always apparent. Researchers should consider that the aforementioned cases could potentially impede a comprehensive understanding and accurate assessment of public opinion regarding mandatory

vaccination.

6.1. Potential applications

Understanding public sentiment toward healthcare policies, institutions, or specific healthcare issues can be crucial for healthcare providers and policy-makers. Stance detection models can help gauge public opinion and adjust strategies accordingly. Another potential application of such systems is disinformation identification. Social media is rife with healthcare misinformation and disinformation. Determination and categorization of content that spreads false or misleading information about healthcare can help authorities respond promptly with accurate information or warnings. In conclusion, the presented data set about vaccine mandates aims to promote research in tracking discussions around vaccines and identify instances of vaccine hesitancy. Healthcare organizations can use this information to design targeted vaccine education campaigns to address specific concerns and increase vaccination rates.

7. Conclusion

The objective of the study was to evaluate various models for stance detection and premise classification on COVID-related tweets. We observed a strong interest in the SMM4H 2022 Task 2, with 47 participants registered, and 14 teams submitted their prediction for both tasks. The models included classifiers on general-domain LMs and LMs, pre-trained specifically on COVID-related tweets, models leveraging syntax features, and dual architectures. We found that models trained on the SMM4H 2022 Task 2 train set of three claims could be successfully transferred to tweets in a new COVID-related claim. As mentioned above, it would be interesting to conduct a more in-depth analysis to study the relationship between emotions and stances on biomedical topics.

Acknowledgements

The work has been supported by the Russian Science Foundation grant # 23-11-00358.

References

- [1] D. Küçük, F. Can, Stance detection: A survey, *ACM Computing Surveys (CSUR)* 53 (2020) 1–37. doi:10.1145/3369026.
- [2] S. Kreps, D. Kriner, Will covid-19 vaccine mandates affect attitudes toward the vaccine and participation in mandate-affected activities? evidence from the united states, *Vaccine* 40. doi:10.1016/j.vaccine.2022.02.083.
- [3] V. Davydova, E. Tutubalina, SMM4H 2022 task 2: Dataset for stance and premise detection in tweets about health mandates related to COVID-19, in: *Proceedings of The Seventh Workshop on Social Media Mining for Health Applications, Workshop & Shared Task, Association for Computational Linguistics, Gyeongju, Republic of Korea, 2022*, pp. 216–220. URL <https://aclanthology.org/2022.smm4h-1.53>
- [4] D. Weissenbacher, J. Banda, V. Davydova, D. Estrada Zavala, L. Gasco Sánchez, Y. Ge, Y. Guo, A. Klein, M. Krallinger, M. Leddin, A. Magge, R. Rodriguez-Esteban, A. Sarker, L. Schmidt, E. Tutubalina, G. Gonzalez-Hernandez, Overview of the seventh social media mining for health applications (#SMM4H) shared tasks at COLING 2022, in: *Proceedings of The Seventh Workshop on Social Media Mining for Health Applications, Workshop & Shared Task, Association for Computational Linguistics, Gyeongju, Republic of Korea, 2022*, pp. 221–241. URL <https://aclanthology.org/2022.smm4h-1.54>
- [5] A. Sakhovskiy, E. Tutubalina, Multimodal model with text and drug embeddings for adverse drug reaction classification, *Journal of Biomedical Informatics* 135 (2022) 104182.
- [6] J. M. Banda, R. Tekumalla, G. Wang, J. Yu, T. Liu, Y. Ding, E. Artemova, E. Tutubalina, G. Chowell, A large-scale covid-19 twitter chatter dataset for open scientific research—an international collaboration, *Epidemiologia* 2 (3) (2021) 315–324.
- [7] A. Wüthrl, R. Klinger, Claim detection in biomedical Twitter posts, in: *Proceedings of the 20th Workshop on Biomedical Language Processing, Association for Computational Linguistics, Online, 2021*, pp. 131–142. doi:10.18653/v1/2021.bionlp-1.15. URL <https://aclanthology.org/2021.bionlp-1.15>
- [8] L. Miao, M. Last, M. Litvak, Twitter data augmentation for monitoring public opinion on COVID-19 intervention measures, in: *Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020, Association for Computational Linguistics, Online, 2020*. doi:10.18653/v1/2020.nlpCOVID19-2.19. URL <https://aclanthology.org/2020.nlpCOVID19-2.19>
- [9] E. Mutlu, T. Oghaz, J. Jasser, E. Tütüncüler, A. Rajabi, A. Tayebi, O. Ozmen, I. Garibay, A stance data set on polarized conversations on twitter about the efficacy of hydroxychloroquine as a treatment for covid-19 (08 2020). doi:10.13140/RG.2.2.34764.03207.

- [10] T. Beck, J.-U. Lee, C. Viehmann, M. Maurer, O. Quiring, I. Gurevych, Investigating label suggestions for opinion mining in German covid-19 social media, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Association for Computational Linguistics, Online, 2021, pp. 1–13. doi:10.18653/v1/2021.acl-long.1
URL <https://aclanthology.org/2021.acl-long.1>
- [11] E. Kotelnikov, N. Loukachevitch, I. Nikishina, A. Panchenko, Ruarg-2022: Argument mining evaluation, Vol. 2022, 2022, p. 333 – 348, cited by: 1; All Open Access, Bronze Open Access, Green Open Access. doi:10.28995/2075-7182-2022-21-333-348.
- [12] L. He, C. He, T. L. Reynolds, Q. Bai, Y. Huang, C. Li, K. Zheng, Y. Chen, Why do people oppose mask wearing? A comprehensive analysis of U.S. tweets during the COVID-19 pandemic, Journal of the American Medical Informatics Association 28 (7) (2021) 1564–1573. arXiv:<https://academic.oup.com/jamia/article-pdf/28/7/1564/38983252/ocab047.pdf>, doi:10.1093/jamia/ocab047.
URL <https://doi.org/10.1093/jamia/ocab047>
- [13] D. Purwitasari, C. B. P. Putra, A. B. Raharjo, A stance dataset with aspect-based sentiment information from indonesian covid-19 vaccination-related tweets, Data in Brief 47 (2023) 108951. doi:<https://doi.org/10.1016/j.dib.2023.108951>.
URL <https://www.sciencedirect.com/science/article/pii/S2352340923000690>
- [14] S. Cheatham, P. Kummervold, L. Parisi, B. Lanfranchi, I. Croci, F. Comunello, M. Rota, A. Filia, A. Tozzi, C. Rizzo, F. Gesualdo, Understanding the vaccine stance of italian tweets and addressing language changes through the covid-19 pandemic: Development and validation of a machine learning model, Frontiers in Public Health 10. doi:10.3389/fpubh.2022.948880.
- [15] H. Alhazmi, Arabic twitter conversation dataset about the covid-19 vaccine, Data 7 (11). doi:10.3390/data7110152.
URL <https://www.mdpi.com/2306-5729/7/11/152>
- [16] K. Glandt, S. Khanal, Y. Li, D. Caragea, C. Caragea, Stance detection in COVID-19 tweets, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Association for Computational Linguistics, Online, 2021, pp. 1596–1611. doi:10.18653/v1/2021.acl-long.127.
URL <https://aclanthology.org/2021.acl-long.127>
- [17] J. Hartmann, Emotion english distilroberta-base, <https://huggingface.co/j-hartmann/emotion-english-distilroberta-base/> (2022).

- [18] P. Ekman, An argument for basic emotions, *Cognition and Emotion* 6 (3-4) (1992) 169–200.
 arXiv:<https://doi.org/10.1080/02699939208411068>, doi:10.1080/02699939208411068.
 URL <https://doi.org/10.1080/02699939208411068>
- [19] F. B. Oliveira, D. Mougouei, A. Haque, J. S. Sichman, H. K. Dam, S. Evans, A. Ghose, M. P. Singh, Beyond fear and anger: A global analysis of emotional response to covid-19 news on twitter using deep learning, *Online Social Networks and Media* (2023) 100253doi:<https://doi.org/10.1016/j.osnem.2023.100253>.
 URL <https://www.sciencedirect.com/science/article/pii/S2468696423000125>
- [20] N. Pivecka, R. Ratzinger, A. Florack, Emotions and virality: Social transmission of political messages on twitter, *Frontiers in Psychology* 13. doi:10.3389/fpsyg.2022.931921.
- [21] J. Schöne, B. Parkinson, A. Goldenberg, Negativity spreads more than positivity on twitter after both positive and negative political situations, *Affective Science* 2 (2021) 1–12. doi:10.1007/s42761-021-00057-7.
- [22] M. Das, A. Mangrulkar, I. Manchanda, M. Kapadnis, S. Patnaik, Enlp musk@SMM4H’22 : Leveraging pre-trained language models for stance and premise classification, in: *Proceedings of The Seventh Workshop on Social Media Mining for Health Applications, Workshop & Shared Task, Association for Computational Linguistics, Gyeongju, Republic of Korea, 2022*, pp. 156–159.
 URL <https://aclanthology.org/2022.smm4h-1.42>
- [23] H. Yang, Z. Zhang, Y. Zhang, yiriyou@SMM4H’22: Stance and premise classification in domain specific tweets with dual-view attention, in: *Proceedings of The Seventh Workshop on Social Media Mining for Health Applications, Workshop & Shared Task, Association for Computational Linguistics, Gyeongju, Republic of Korea, 2022*, pp. 23–26.
 URL <https://aclanthology.org/2022.smm4h-1.7>
- [24] S. Chen, Q. Jin, Multi-modal conditional attention fusion for dimensional emotion prediction, in: *Proceedings of the 24th ACM International Conference on Multimedia, MM ’16, Association for Computing Machinery, New York, NY, USA, 2016*, p. 571–575. doi:10.1145/2964284.2967286.
 URL <https://doi.org/10.1145/2964284.2967286>
- [25] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, I. Polosukhin, Attention is all you need, in: I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), *Advances in Neural Information Processing Systems*, Vol. 30, Curran Associates, Inc., 2017.
 URL https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf
- [26] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, cite

arxiv:1810.04805Comment: 13 pages (2018).

URL <http://arxiv.org/abs/1810.04805>

- [27] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, L. Zettlemoyer, BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online, 2020, pp. 7871–7880. doi:10.18653/v1/2020.acl-main.703. URL <https://aclanthology.org/2020.acl-main.703>
- [28] M. Honnibal, I. Montani, spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing, to appear (2017).
- [29] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, D. Krishnan, Supervised contrastive learning, in: H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, H. Lin (Eds.), Advances in Neural Information Processing Systems, Vol. 33, Curran Associates, Inc., 2020, pp. 18661–18673. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/d89a66c7c80a29b1bdbab0f2a1a94af8-Paper.pdf
- [30] M. Müller, M. Salathé, P. E. Kummervold, COVID-Twitter-BERT: A Natural Language Processing Model to Analyse COVID-19 Content on Twitter, arXiv preprint arXiv:2005.07503.
- [31] I. Loshchilov, F. Hutter, Decoupled weight decay regularization, in: International Conference on Learning Representations, 2019. URL <https://openreview.net/forum?id=Bkg6RiCqY7>
- [32] C. Xu, C. Paris, S. Nepal, R. Sparks, C. Long, Y. Wang, Dan: Dual-view representation learning for adapting stance classifiers to new domains, in: ECAI 2020, IOS Press, 2020, pp. 2260–2267.
- [33] I. Loshchilov, F. Hutter, Decoupled weight decay regularization, in: International Conference on Learning Representations, 2017.
- [34] P. Refaeilzadeh, L. Tang, H. Liu, Cross-validation, Encyclopedia of database systems 5 (2009) 532–538.
- [35] F. Wilcoxon, Individual comparisons by ranking methods, Biometrics Bulletin 1 (6) (1945) 80–83. URL <http://www.jstor.org/stable/3001968>

Appendix

Tweet text	Emotion
<i>Santa popped by one of our south west London vaccination centres to say thank you to all the wonderful vaccinators and volunteers! Make yourself a present by getting your Covid vaccination for Christmas!</i>	joy
<i>Believe in God! Protect yourself from these satan worshippers injecting this poison into you! This is not a vaccine, but Trojan horse to introduce nanotechnology which will be integrated with travel, Commerce and digital currency etc. They want to kill billions by 2025!</i>	fear
<i>They've already begun in Quebec "Vaccination" of the 5-11yr olds started around December 6th (date varied on schools/sectors) These sick! Just before Christmas too..Absolutely disgusting! #LeaveOurKidsAlone #NoMandatoryVaccines</i>	anger
<i>Very caring and sensible of you. Our Xmas is cancelled too, as I'm high risk for both Covid and the vax, tho thankfully managed to get my 1st, in hospital earlier this week, without too many issues. But not risking my being out about over the Christmas break. Stay safe over there, darling.</i>	love
<i>Everyone is tired and exhausted. It's made much worse when you don't see the light at the end of the tunnel. For so long that vaccination rate was the light. We'd be ok when we got there, now what? What's keeping the hope alive now? The next booster?</i>	sadness
<i>Wow, I'm shocked! Maybe they should reconsider the vax mandates.</i>	surprise

Table .6: Examples of tweets about COVID-19 vaccines and corresponding emotions.

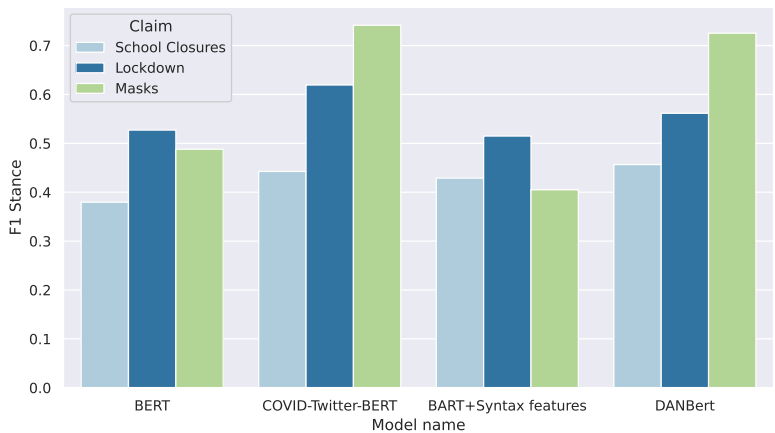


Figure .4: F1 metrics for stance detection on 3 claims: School closures, Stay-at-home orders. All models were trained on tweets about these claims.

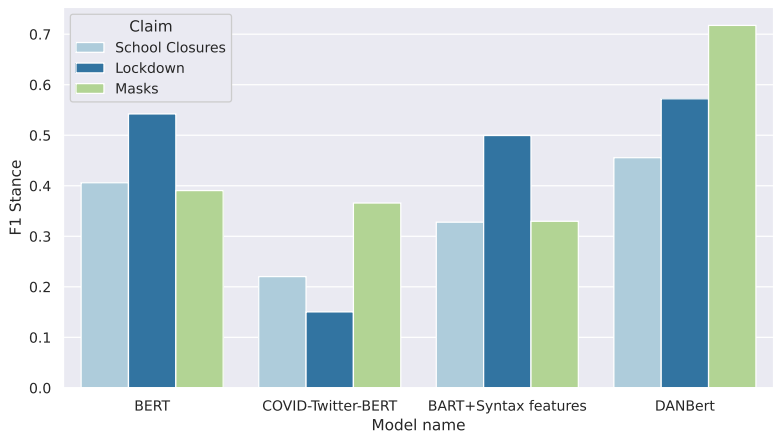


Figure .5: F1 metrics for stance detection on 3 claims: School closures, Stay-at-home orders. All models were trained on tweets and corresponding claims.

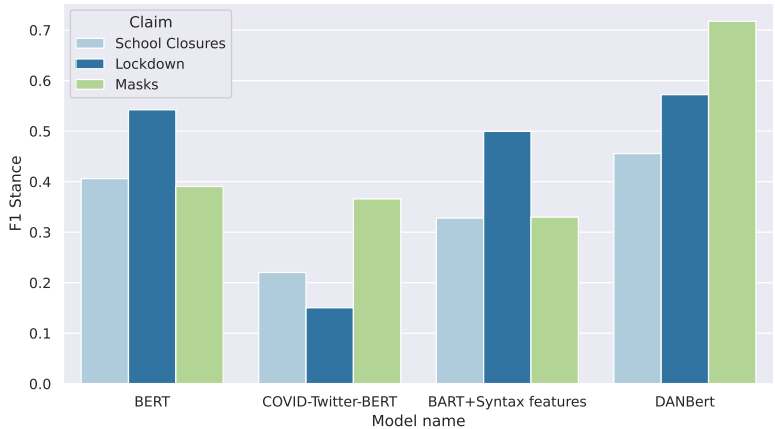


Figure .6: F1 metrics for premise detection on 3 claims: School closures, Stay-at-home orders. All models were trained on tweets about these claims.

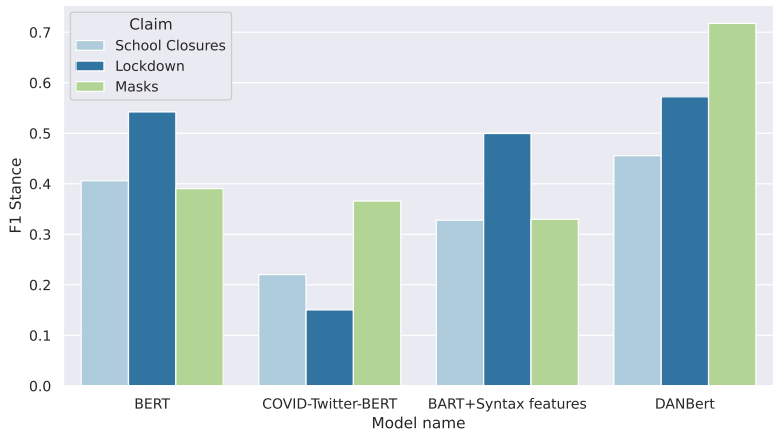


Figure .7: F1 metrics for premise detection on 3 claims: School closures, Stay-at-home orders. All models were trained on tweets and corresponding claims.

Hashtags	Stance
GetVaxxed VaccinationWorks getvaccinated VaxxedSquad MoreVaccinesMoreLivesSaved vaccinesafetyadvocate justwantmyjab IChooseVaccination VaccinesSaveLives vaccineworks getvaccinated YourBodyYourChoice GetVaxed GetVaccinatedX3 vaccinesavelifes GET-VACCINATED VaccineMandatesNOW vaccineswork vaccinatethe-world GetVaccinatedASAP Vaccine4All VaccinateTheWorld GetVaccinatedNow vaccinateASAP vaccinateyourself VaccinesWork getvaccinatedtobeprotected Vaccinate4All GetVax vaccinessaveslives iChooseVaccination VaxToTheMax GetVaccinated	favor
NoVaccine novaccinemandates Vaccinationisachoice NoVaccineMandates NoVaxPass NoVaccineMandate vaccinedeaths vaccineinjured UNVACCINATED NoVaccinePassportsAnywhere VaccineDeaths mybodymychoice vaccineinjury pharmacide NoVaccinePassports EndVaccineMandatesNow C19vaxKills NoVax NoVaccinePassport NoMandatoryVaccines NoVaccinePassportAnywhere vaccineinjuries NoVaccine_NoPandemic jabskill NoVaccineMandatesAnywhere saynotoVaccinemandate prochoice NoVaccinePassportAnywhere	against

Table .7: List of hashtags related to vaccine mandates used for tweets extraction and pre-annotation of tweets. Individual hashtags are separated by spaces.