Journal of Scientific & Industrial Research
Vol. 83, April 2024, pp. 381-391
DOI: 10.56042/jsir.v83i4.1012

Stance and Sentiment Analysis of Health-related Tweets with Data Augmentation

Doğan Küçük[1]* & Nursal Arıcı[2]

[1]Department of Computer Engineering, Graduate School of Natural and Applied Sciences, Gazi University, Turkey

[2]Department of Management Information Systems, Faculty of Applied Sciences, Gazi University, Turkey

Received 29 April 2023; revised 10 November 2023; accepted 20 February 2024

Common social media platforms like Twitter are important as up-to-date information sources for several monitoring purposes, including instant public health monitoring. In this sense, large volumes of health-related social media posts (such as tweets on the COVID-19 pandemic) have been produced recently, and are ready to be analyzed to facilitate health-related decision making. In this paper, joint stance detection and sentiment analysis on tweets about the COVID-19 vaccination was performed, in order to showcase the contribution of different machine learning and deep learning techniques equipped with data augmentation. Training and test tweet datasets are compiled and annotated for both stance and sentiment analysis and next, the training dataset is extended using an automatic data augmentation technique to increase its size. Experiments with different classifiers are performed for automated stance and sentiment analyses, using this extended dataset during training. The data augmentation technique adopted in this study to cope with data scarcity problems in machine learning research leads to better performance rates in this domain of health-related social media analysis. Comparative evaluations are also performed using a publicly-available sentiment analysis tool. The extended dataset and the test dataset, along with the approaches, and evaluation results are significant for health informatics, because, they facilitate joint estimation of instant community stance and sentiment towards COVID-19 vaccination which has been an important public health concern. Therefore, public health decision-makers can extensively and readily benefit from the findings and resources of the current study.

Keywords: Deep learning, Health informatics, Machine learning, Natural language processing, Public health monitoring

Introduction

Public health monitoring and surveillance are usually performed with electronic health records of hospitals, as these records are formal sources of health information. Surveys have also been used commonly to collect common health-related data. But, conducting surveys is a labor-intensive and requires considerable time to obtain the results. Analyzing social media posts is another plausible choice, which enables automatic collection and monitoring of health-related data.[1] But this analysis should be conducted cautiously to avoid misinformation and disinformation that can quickly spread via social media.[2]

Especially after the global COVID-19 pandemic, many studies target COVID-19-related social media posts. For instance, some studies publish large datasets including COVID-19-related posts.[3,4] By analyzing these texts, researchers aim to extract the opinions of the populations about the pandemic and other related topics.[5] These related topics include COVID-19 vaccines and drugs, using face masks, in addition to lockdowns and quarantines to prevent the spread of the disease.

Some studies perform sentiment analysis and emotion recognition on the collected COVID-19 related social media texts. Generally, sentiment analysis (opinion mining) aims to classify input text as *positive*, *negative*, or *none*.[6,7] In emotion recognition, the emotion expressed in the text is explored, like *joy*, *fear*, *sadness*, *anger*, etc.[8] Other related studies perform stance analysis (stance detection) towards controversial issues (vaccination, quarantines, etc.) related to the pandemic. In stance detection, the position (as *favor*, *against*, and *none* (or, *neither*)) of the text author towards a particular target or a set of targets is determined.[9,10]

This study aims to apply joint stance detection and sentiment analysis on tweets about COVID-19 vaccination, to showcase the contribution of different machine learning techniques equipped with data augmentation, to extract affective information from health-related social media posts automatically.

Current study started by downloading a tweet dataset about this target topic and annotating the

*Author for Correspondence
E-mail: dogan.kucuk@gazi.edu.tr

dataset with sentiment and stance labels, COVID-19 vaccination being the stance target. A *back-translation* based data augmentation technique was used in this study to double the size of the annotated dataset automatically. Finally, both stance detection and sentiment analysis was performed using SVM and logistic regression models in addition to bagging and random forest algorithms, where the latter two are ensemble models. The performance of a deep learning model, namely Long Short-Term Memory (LSTM) was tested. The performance rates of these machine and deep learning models were presented and the sentiment analysis results of this study were compared with the results of a publicly-available sentiment analyzer for social media.

The main contributions of this paper are as below:

1) Social media analysis for health-related information extraction is an important research topic. Related research results may have practical contributions to the decision-making procedures during pandemics such as the current one. The topic of this paper, namely COVID-19 vaccination, is a controversial public health issue. Applying stance detection and sentiment analysis on social media posts about COVID-19 vaccination may contribute to a better understanding of the concerns and attitudes of people related to this recent and important public health issue by public health professionals.
2) Stance detection and sentiment analysis are important and popular problems of natural language processing, particularly on social media. These significant social media analysis techniques can readily be used to discover or estimate the public opinions and positions/stances of citizens, which can be analyzed for improved public health decision-making.
3) Data scarcity is the lack of (mostly annotated) data to be used by learning algorithms, and *data augmentation* is an important attempt to cope with data scarcity. It is increasingly used in several research domains related to natural language, speech, and image processing. This study used a translation-based data augmentation technique, called *back-translation*, to automatically extend the initial annotated tweet dataset.
4) This study have employed and tested four machine learning models and a deep learning model during experiments. The performance of these models was compared and the results were discussed. For the sentiment analysis task, the performance of a publicly available sentiment analyzer was tested and its results were compared with that of the five learning models used.
5) Although sentiment analysis is well-studied, stance detection is a new research area, especially for low-resource languages like Turkish. This study is significant for its jointly performing sentiment analysis and stance detection on health-related social media posts. Another contribution of the study is its creation of linguistic resources for Turkish since related resources for the Turkish language are still limited compared to languages like English, Spanish etc. The training and test datasets in Turkish that are annotated for stance and sentiment analysis can help increase the breadth and depth of related future studies.

Related Work

Beginning with the outbreak of the COVID-19 pandemic, several related social media analysis studies have been performed. An important share of these studies is on datasets of social media posts related to COVID-19. Monolingual or multilingual tweet datasets about the COVID-19 pandemic are described in several studies.[3,4,11,12]

Apart from the studies that present COVID-19-related datasets, there are recent studies on sentiment, stance, and/or emotion analysis on COVID-19-related social media posts, such as the recent relevant study.[13] Since the current paper focuses on sentiment and stance analysis towards COVID-19 vaccination, these studies are particularly considered and reviewed in the rest of this section.

Sentiment analysis results on Twitter and Facebook posts on COVID-19 vaccines (from the UK and US) are presented in related studies where sentiment analysis is performed using BERT (Bidirectional Encoder Representations from Transformers) language model.[5,14] A project that aims to monitor Twitter and Facebook posts related to COVID-19 vaccines in Italy is described.[15] The concept drift problem in machine learning algorithms to determine vaccine sentiment during the pandemic is addressed.[16] A sentiment analysis approach on COVID-19-related tweets using BERT is described.[17] Topic modeling, sentiment and emotion analysis are performed to detect public perceptions of Twitter users towards COVID-19 vaccines.[18] English and Italian tweets on COVID-19 vaccines are analyzed to determine community perceptions of COVID-19 vaccine campaigns.[19]

The focus of another related study is rumor stance detection on tweets about COVID-19 to determine the spread of misinformation on Twitter.[20] Rumor stance detection is a problem related to two issues: stance detection and rumor detection. Given a piece of text and a rumor as input, the position in the input text towards the input rumor is expected at the end of the rumor stance detection procedure.[9] Therefore, the rumor classes (*endorse*, *deny*, *question*, and *neutral*) usually differ from the generic stance classes.[20]

The sentiments of Australian Twitter users towards COVID-19 vaccination are analyzed, in addition to topic modeling and emotion analysis on the same tweet dataset.[21] Sentiment analysis and topic modeling are applied to tweets about COVID-19 vaccines in another related study.[22] Social media analysis is utilized to determine the attitudes and behaviors of people related to COVID-19 vaccine trials.[23] A tweet analysis process is performed to determine the opinions and sentiments of the Japanese population towards COVID-19 vaccination.[24] Another related study aims to determine the evolution of Twitter threads regarding influenza vaccination during the pandemic.[25] The authors of a related work perform k-means clustering during opinion mining on Twitter about COVID-19 vaccination.[26] Topic modeling is used to characterize tweets regarding COVID-19 vaccines in the US.[27] Similarly, sentiment analysis is applied to tweets from the US about the COVID-19 vaccines in several studies.[28,29]

A systematic review of studies concerning public attitudes on social media platforms towards COVID-19 vaccination has recently been presented.[30] It is concluded that Twitter is used as the social media source in a vast percentage of the related studies and that social media can be a critical means to prevent related public health issues such as vaccine hesitancy.[30] Another related literature survey covers those studies that perform sentiment, emotion, and stance analysis on tweets about COVID-19 vaccination.[31] It is similarly emphasized that the findings of the relevant papers can help health-related policymakers improve public health during pandemics.[31]

Different traditional machine learning algorithms are evaluated for sentiment analysis on tweets about COVID-19 vaccines.[32] Linear SVC (a support vector classification model) with features based on TF-IDF is reported to be the best-performing learning model for determining sentiment (*positive*, *negative*, *neutral*) on this dataset.[32]

Tweets published in India about the side effects of COVID-19 vaccines are collected and the sentiments of the tweet producers are determined using the TextBlob sentiment analyzer.[33] The results reveal that more than 78% of the tweets are either *positive* or *neutral*, which can be considered a favorable outcome to address related issues such as vaccine hesitancy in India.[33]

Tweets about COVID-19 vaccines are analyzed for sentiment using the publicly-available VADER rule-based sentiment analysis tool.[34,35] Almost 47% of the tweets are classified as *positive* or *highly positive*, 24% as *negative* or *highly negative*, and 29% as *neutral*. While the tweets with *positive* sentiments are more about getting vaccination, those with *negative* sentiments are mostly on vaccine hesitancy and its side effects.[35] Sentiment analysis is applied to a Kaggle dataset of tweets about COVID-19 vaccination from different countries, again with the VADER sentiment analyzer.[36] The results of this study indicate that *positive* sentiment is more prevalent in the analyzed tweets than *neutral* and *negative* sentiments.[36]

To sum up, there are recent papers regarding (commonly) sentiment analysis and (rarely) stance detection of social media posts about COVID-19 vaccination/vaccines. Most of the related work is performed on Twitter, and targeting a single text analytics problem at a time (such as sentiment analysis), only.

The current study contributes to the related work with the compiled health-related datasets comprising stance and sentiment-annotated tweets, by jointly performing sentiment analysis and stance detection using different machine learning models, and by employing a plausible data augmentation technique to deal with the data scarcity problem that leads to improved performance rates for both problems.

Approaches and Datasets

Proposed Approach

The flowchart demonstrating the approach for data augmentation and subsequent stance and sentiment analysis experiments on tweets about COVID-19 vaccination is given in Fig. 1.

Since the details of the phases given in this flowchart are included in the upcoming subsections of the paper, a summary of the overall procedure is given as follows:

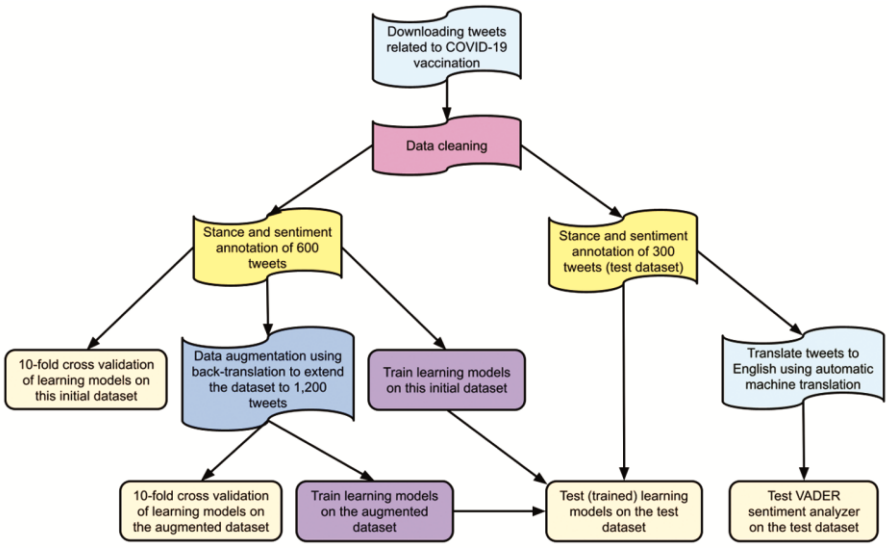

Fig. 1 — The flowchart of the proposed approach

- This study started by downloading tweets using the Twitter API, and then applying data cleaning to filter out duplicate tweets, tweets in other languages, and retweets. The first part of the data collection phase includes the annotation of 600 tweets (training dataset) using stance and sentiment labels. This study performed stance and sentiment experiments using a special set of machine learning models on this initial dataset with 10-fold cross-validation.
- Then, data augmentation technique based on back-translation was applied to automatically double the size of this annotated dataset and obtain the augmented dataset. Similarly, machine learning experiments were carried on the augmented dataset with 10-fold cross-validation.
- On the other hand, similarly another dataset of 300 tweets about COVID-19 vaccination was compiled, cleaned, and annotated, to be used as an independent test dataset. The learning models were tested on this new test dataset after training.
- The commonly available VADER sentiment analyzer was also evaluated on the test dataset, to compare its performance with the sentiment analysis performance of the learning models employed.
- Finally evaluation results of this study were compared under these different settings and the obtained findings were discussed.

Initial Tweet Dataset on COVID-19 Vaccination

First 600 tweets in Turkish on COVID-19 vaccination were collected through the Twitter API (https://developer.twitter.com/en/docs/twitter-api/tweets/search/introduction) where Tweepy (https://www.tweepy.org/) open-source Python library is utilized to access the API. To ensure the balance of opinions, each of the two halves (300 tweets each) of the tweets belongs to different dates. There is a period of more than six months between the dates of the two halves of the dataset.

The word cloud of the tweets in this initial dataset based on highly-frequent words in the tweets is provided in Fig. 2.

The most frequent (top-3) words in the dataset are "aşı" (meaning "*vaccine*"), "*saglikbakanligi*" (hashtag-like use of "*SağlıkBakanlığı*" (meaning "*Ministry of Health*"), and "*drfahrettinkoca*" (Twitter username of Turkish Minister of Health).

These tweets are first annotated with stance classes (*favor*, *against*, and *none*) using COVID-19 vaccines as the stance target. Then, they are also annotated with the sentiment class labels of *positive*, *negative*, and *none*. Regarding stance; 259 tweets are marked as *favor*, 245 as *against*, and 96 as *none*. In terms of sentiment; 44 tweets are marked as *positive*, 285 as *negative*, and 271 as *none*.

The distribution of these class labels for sentiment analysis and stance detection/classification is also shown in Fig. 3. This preliminary work presents the dataset of this study and some machine learning tests were carried out on this dataset.[37]

The annotations are carried out following the stance and sentiment annotation guidelines.[10] Considering the result of stance annotations, the numbers of tweets marked as *favor* and *against* are

Fig. 2 — Word cloud of the first tweet dataset

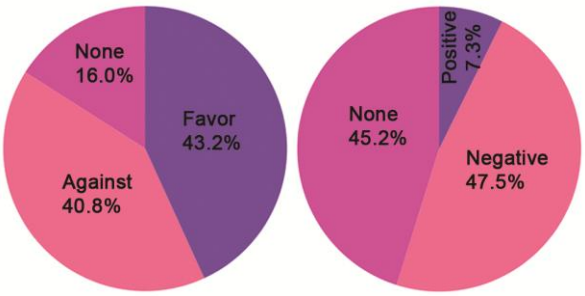

Fig. 3 — The distribution of stance and sentiment class labels in the first dataset

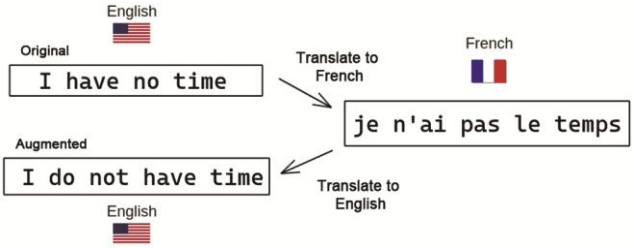

Fig. 4 — An example showcasing the back-translation data augmentation technique (https://amitness.com/2020/02/back-translation-in-googlesheets/)

almost equal, and are more frequent than those tweets marked as *none*.

Yet, with respect to sentiment annotations on the dataset, the numbers of tweets marked as *negative* or as *none* are more frequent and quite close to each other while the number of tweets marked as *positive* is very low. This lack of positive opinions in the dataset can be attributed to the controversial nature of the "*COVID-19 vaccination*" topic: when using Twitter, both the proponents and the opponents of vaccination avoid a positive attitude when publishing their ideas on vaccines, through social media platforms.

Data Augmentation

Data augmentation techniques are used to automatically increase the annotated instances in datasets (to be used by learning algorithms). That is, data augmentation can help address the data scarcity problem in machine learning and deep learning research.

Data augmentation techniques include two-phase translation of sentences to obtain paraphrased sentences.[38] For example, a sentence in English is first translated into French and then back to English.[38] This data augmentation technique is also known as *back-translation* and a concrete example is shown in Fig. 4.

Other basic techniques, such as random swap and synonym replacement, are used for text classification tasks.[39] Data augmentation has also been employed to improve sentiment analysis as well as stance detection.[40,41]

To extend the initial dataset of this study described previously, the machine translation-based data augmentation technique (as also demonstrated in Fig. 4) was used.[39] All of the tweets in the dataset were translated into English, and then back into Turkish, using Google Translate (https://translate.google.com/) as the machine translation tool. At the end of this process, an annotated dataset of

1,200 tweets were obtained, where 600 new items are paraphrased (translated and then back-translated) versions of the original 600 tweets.

Second Tweet Dataset on COVID-19 Vaccination

In order to employ it during experiments as an independent test dataset, another brand-new tweet dataset on COVID-19 vaccination was compiled and annotated. This test dataset includes 300 tweets (published on a different date) and the word cloud of these tweets is displayed in Fig. 5. The most common three words in this test dataset are "*aşı*" (meaning "*vaccine*"), "*zorunlu*" (meaning "*compulsory*"), and "*yok*" (meaning "*does not exist*").

This second dataset is also annotated with stance and sentiment classes, and at the end of the annotation process, there are 108 tweets marked as *favor*, 121 labeled as *against*, and 71 labeled as *none* (in terms of stance). There are 32 tweets marked as *positive*, 167 marked as *negative*, and 101 marked as *none* (in terms of sentiment). The percentages of the class labels in this second dataset are displayed in Fig. 6.

Similar to the previous observation on the extended dataset, the number of tweets with *positive* sentiment is also relatively low in this dataset. Again, after examining the tweets in this dataset, it can be claimed that both the proponents and the opponents of COVID-19 vaccination tend to use *negative* or *neutral* language instead of *positive* one.

Evaluation Results

Metrics

Precision, *Recall*, and *F-score* (also known as *F1-score* or *F-measure*) are among the most common evaluation metrics used in sentiment analysis and stance detection studies. The formulae used to calculate Precision, Recall, and eventually (their harmonic mean) F-score are provided below. Due to its widespread employment in the related literature as a single metric calculated from precision and recall, the F-score metric was utilized during the stance and sentiment analysis experiments.

$$\text{Precision (\%)} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \times 100$$

$$\text{Recall (\%)} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \times 100$$

$$\text{F} - \text{Score (\%)} = \frac{2 \text{ x Precision x Recall}}{\text{Precision} + \text{Recall}} \times 100$$

Employed Learning Approaches

The following five machine learning algorithms were used during the stance and sentiment analysis experiments on the health-related tweet datasets compiled and extended (all from the Weka learning toolkit):[42]

- *Bagging* belongs to the category of *ensemble learners* which utilize more than one classifier and combine the results of these classifiers to produce a better output. Although it uses more than one classifier, all of the classifiers belong to

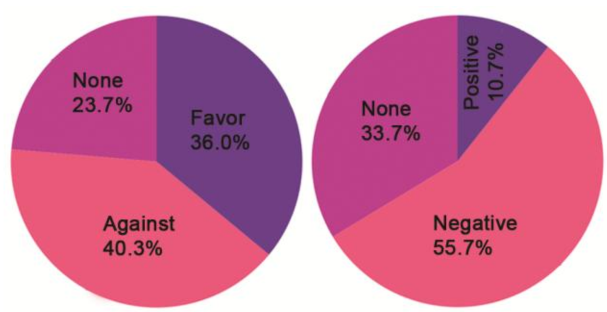

Fig. 6 — The distribution of stance and sentiment class labels in the second dataset

Fig. 5 — Word cloud of the second tweet dataset

the same category in the bagging ensemble classifier, and in this case, bagging uses a decision tree classifier. Hence, this ensemble classifier is also referred to as *bagged decision trees* in the related literature.

- *Logistic Regression* is referred to as a *statistical classification model* that can be used to predict a categorial outcome, which has been reported achieve high performance for text processing tasks including stance detection.[9] Logistic regression differs from linear regression which predicts a continuous outcome, as the former one is a supervised classification model that outputs categorial labels.
- *SVM* is a traditional machine learning algorithm based on the concept of optimized *hyperplane separation* and it has widespread use in different application domains including various text analysis tasks such as stance detection and sentiment analysis. During the experiments, polynomial kernel is used by the employed SVM implementation.
- *Random Forest* is another *ensemble learning model*, similar to the bagging model described above. Random forest also aggregates the results of individual decision tree classifiers in order to output the final class labels. It differs from bagged decision trees mainly in that random subsets of data and attributes are used to determine the split points of the underlying decision trees.
- *LSTM* is a neural network-based *deep learning model* and is an instance of recurrent neural networks (RNNs).[43] LSTMs are known to learn long-range patterns successfully, as an advantage over RNNs. LSTMs have been applied to different natural language processing tasks and hence, the current settings have employed the LSTM model for stance and sentiment analysis by using the WekaDeeplearning4j package installed on Weka learning toolkit.[44]

Stance and Sentiment Analysis Experiments

In the preliminary work, stance and sentiment analysis was performed using SVM and random forest classifiers.[37] Unigrams, the existence of hashtags, and the existence of emoticons were used as the features of these classifiers. Tests were conducted by performing 10-fold cross-validation on the initial dataset (of 600 tweets) only. The performance rates of SVM classifiers were reported to be higher than that of random forest classifiers.[37]

As pointed out in the previous sections, extending this initial dataset by data augmentation, compiling and annotating another tweet dataset for testing purposes, and performing experiments with three new learning models (bagging, logistic regression, and LSTM) are among the novel aspects of the current paper. Another novel aspect is using sentiment classes as an additional feature during stance analysis, and similarly, using stance classes as an additional feature during sentiment analysis.

This study performed various stance and sentiment analysis experiments using the five classifiers described in the previous subsection, under different settings. The experiments are carried on a MacOS computer with 8GB of RAM and an ARM-based M1 chip, using the Weka learning toolkit (version 3.8.6).[(42)]

First of all, the stance and sentiment analysis performance rates of the five classifiers were obtained on the initial and augmented (extended) datasets, using 10-fold cross validation as demonstrated in Table 1.

In Table 1, the highest F-scores are shown in boldface. It can be observed from these experimental results on the initial dataset that SVM surpasses the other classifiers in F-score for the stance detection task; however, the performance rates of SVM, bagging, LSTM, and logistic regression are close to each other for the sentiment analysis task where logistic regression classifier slightly surpasses the other four classifiers. LSTM model and bagged decision tree (bagging) classifier both achieve a very close F-score rate. Although the performance rates of

Table 1 — Evaluation results of sentiment and stance analysis classifiers on the initial and augmented health-related tweet datasets (with 10-fold cross-validation and in F-score)

	Initial Dataset (600 tweets)		Extended Dataset (1,200 tweets)	
Classifier	Stance Analysis	Sentiment Analysis	Stance Analysis	SentimentAnalysis
Bagging	54.6%	62.0%	61.8%	67.0%
Logistic Regression	51.4%	**62.7%**	56.7%	66.0%
SVM	**61.5%**	61.2%	**89.3%**	**88.3%**
Random Forest	57.6%	50.0%	86.6%	76.5%
LSTM	55.9%	62.0%	60.5%	71.9%

random forest correspond to the second-best results for the stance detection problem, random forest falls short of the other models for the sentiment analysis task on the initial dataset.

The results in Table 1 also reveal that logistic regression consistently performs poorly for the stance detection task, and similarly, random forest performs poorly for the sentiment analysis task in the settings where health-related tweets are considered. Hence, the results suggest that the remaining classifiers seem more viable in a practical decision support system for public health surveillance where community stance and sentiment are monitored through social media posts.

In the second experimental settings, the second (separate) dataset of 300 tweets was used as the test set, and use the initial and extended datasets as training datasets, respectively, to observe the contribution of the data augmentation procedure to stance and sentiment analysis tasks on health-related tweets. The corresponding evaluation results in F-score are presented in Table 2 and the best performance rates are highlighted.

The results in Table 2 demonstrate that better sentiment and stance analysis results were obtained by automatically extending the training dataset through data augmentation. SVM classifier is the top performer for stance detection and gains more than 2% improvement in F-score through data augmentation for both problems.

This finding is in line with related work as it has also been reported in related surveys that SVM classifiers usually achieve favorable results for the stance detection problem.[9]

There are many high-impact stance detection studies in the related literature reporting that an SVM-based model surpasses many other learning models used for comparison.[10] For instance, SVM classifiers for stance detection attains F-scores ranging from 42.4% to 66.4% for different stance targets given in a benchmark tweet dataset.[10,45]

On the other hand, similar to the findings in Table 1, the logistic regression classifier surpasses the other classifiers for the sentiment analysis task, although its performance improvement through data augmentation falls behind the other classifiers for both tasks.

As elaborated above, according to the results given in Table 2, SVM is the best performing model for the stance detection task and logistic regression is the best performer for the sentiment analysis task, under test settings of this study. Yet, it should also be noted that the corresponding F-score rates of these classifiers suggest that (as part of future work) further performance improvements are necessary before employing these two classifiers in real-time settings.

Although random forest is the classifier that benefits from data augmentation the most, its overall performance is poor for both problems. For all cases, the F-score rates of the random forest classifier are all below 50% while the other tree-based ensemble classifier, bagging, achieves the second-best results for both problems on the test dataset (with the exception of the last column in Table 2 where LSTM attains the second-best result), and significantly benefits from the data augmentation technique. Particularly for sentiment analysis, bagging achieves performance rates very close to that of logistic regression.

The performance of the deep learning model (LSTM) is promising in general for all four settings given in Table 2. LSTM performs better for the sentiment analysis task as it achieves the second-best F-score for sentiment analysis on the test dataset after being trained on the automatically extended training dataset. However, LSTM is not the top-performer for any of the experimental settings. This observation could be attributed to the scarcity of annotated data for this deep learning model, although it has extended the size of the initial annotated dataset automatically through data augmentation. Hence, it is believed that further experiments on larger extended datasets could be performed to better analyze the performance of

Table 2 — Evaluation results of classifiers for sentiment and stance analysis on the second health-related tweet dataset, using initial and extended datasets for training (in F-score)

	Initial Dataset (training) Second Dataset (testing)		Extended Dataset (training) Second Dataset (testing)	
Classifier	Stance Analysis	Sentiment Analysis	Stance Analysis	Sentiment Analysis
Bagging	44.7%	57.5%	50.4%	59.0%
Logistic Regression	44.5%	**59.7%**	45.8%	**59.8%**
SVM	**51.1%**	55.2%	**53.0%**	57.7%
Random Forest	40.1%	35.1%	45.8%	47.7%
LSTM	43.5%	57.0%	48.7%	59.1%

deep learning models in the problem settings of this study.

As mentioned previously, it should be noted that in a high-impact stance and sentiment analysis study on English tweets, it is reported that an SVM-based classifier achieves stance detection performance rates ranging from 42.4% to 66.4% in F-score for different stance targets, and sentiment analysis performance ranging from 72.6% to 80.7% in F-score.[10] The results are in line with this previous study in the sense that stance detection results are consistently lower when compared to sentiment analysis results and that SVM is a viable classifier for both tasks. Polynomial kernel is used in the SVM experiments, yet, there is a plan to perform hyperparameter tuning to improve the performance of this SVM classifiers, as part of future work.

Comparison

For comparison purposes, VADER, a publicly-available sentiment analysis tool was used.[35] VADER is a commonly-utilized rule-based sentiment analyzer that is tailored to social media posts like tweets.[35]

The test dataset (of 300 tweets) is first automatically translated to English, and next; the VADER sentiment analyzer is executed on these translations. In Table 3, the F-score of VADER is presented and compared to the corresponding sentiment analysis results of the five classifiers, as excerpted from the last column of Table 2.

It can be observed that the performance of VADER is poorer compared to logistic regression, bagging, SVM, and LSTM classifiers trained on the augmented tweet dataset. Yet, VADER performs better than the random forest model. Although one of the reasons of the poor performance of VADER is the machine translation stage introduced before actual sentiment analysis, this low performance can also be an indicator that sentiment analysis is a harder task on health-related tweets, particularly about this controversial topic of COVID-19 vaccination where proponents and opponents frequently use different linguistic phenomena like irony in their social media posts.

Table 3 — Evaluation results of sentiment analysis using the VADER tool and the employed five classifiers on the test dataset (in F-score)

	Sentiment Analysis
VADER	52.3%
Bagging	59.0%
Logistic Regression	**59.8%**
SVM	57.7%
Random Forest	47.7%
LSTM	59.1%

Discussion of the Results

Based on the findings, the following conclusions can be drawn:

- First of all, it can be noted that stance annotations in the training and test tweet datasets are almost balanced with respect to the three stance classes; however, they are not balanced considering the sentiment classes. The number of tweets marked with *negative* and *none* sentiment labels are higher than the number of *positive* tweets. It can be attributed to the observation on the datasets of this study that regardless of their being *in favor* or *against* COVID-19 vaccination, Twitter users are inclined to demonstrate a *non-positive* attitude in their social media posts.
- Data augmentation techniques are useful for addressing the data scarcity problem. In the tests, it was observed that using an automatically extended dataset (with the back-translation data augmentation technique) as the training dataset leads to improved performance rates in F-score for all of the five classifiers employed. Data augmentation can also be considered as a useful technique to alleviate the effects of class imbalance issue, by increasing the size of the dataset. Besides, there is a plan to utilize other techniques to deal with the class imbalance issue in future studies based on the current work.
- Among these classifiers, SVM achieves the best performance rates for the stance detection task and logistic regression achieves the best rates for sentiment analysis on Turkish tweets about COVID-19 vaccination. The bagging classifier consistently attains the second-best rates for both problems, making it a viable alternative for both problems in a practical monitoring and decision support system for public health professionals. The deep leaning model, LSTM, achieves promising results for both problems and its performance for the sentiment analysis task is higher and close to that of the top-performing classifiers. Finally, although it benefits from data augmentation extensively, random forest demonstrates the lowest performance rates for both problems and its performance is even poorer than the VADER sentiment analyzer used for comparison while the former four classifiers all perform better than VADER.
- To sum up, the findings suggest that, among five classifiers, SVM (together with data augmentation

to cope with scarcity) is the best classifier for stance detection towards health-related controversial topics on Twitter, and logistic regression, bagging, and LSTM models (with data augmentation) are more convenient for sentiment analysis on health-related social media posts.

- As part of future work, stance and sentiment experiments with other deep learning models will be performed.[46] Other natural language processing problems such as named entity recognition can also be included within the scope of the further experiments on the tweet datasets.[47]

Conclusions

This study performed joint sentiment analysis and stance detection on tweets about COVID-19 vaccination to automatically obtain the opinions and positions of people, regarding this controversial topic. Based on the experiments with different machine and deep learning models including bagged tree, logistic regression, SVM, random forest, and LSTM, it is observed that increasing the training dataset with an automatic data augmentation scheme improves the performance of sentiment and stance analysis with all of the classifiers. The findings of the experiments also suggest that SVM classifiers are more promising for stance detection while logistic regression and bagging classifiers are more convenient for sentiment analysis under experimental settings. Further studies include testing other data augmentation techniques, learning models, and feature sets, and using the best performing models and attributes during the implementation of a public health surveillance and decision support system for public health professionals and decision-makers, based on the analysis of social media posts for instant public health monitoring.

References

1 Küçük D, Arıcı N & Küçük E E, Automatic public health monitoring on social media: A recent survey, *Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi*, **10(2) (**2021) 439–449.

2 Conforti C, Berndt J, Pilehvar M T, Giannitsarou C, Toxvaerd F & Collier N, Will-they-won't-they: A very large dataset for stance detection on Twitter, *arXiv preprint arXiv:2005.00388* (2020).

3 Haouari F, Hasanain M, Suwaileh R & Elsayed T, Arcov-19: The first Arabic COVID-19 Twitter dataset with propagation networks, *arXiv preprint arXiv:2004.05861* (2020).

4 Lamsal R, Design and analysis of a large-scale COVID-19 tweets dataset, *Appl Intell*, **51 (**2021) 2790–2804, https://doi.org/10.1007/s10489-020-02029-z.

5 Hussain A, Tahir A, Hussain Z, Sheikh Z, Gogate M, Dashtipour K, Ali A & Sheikh A, Artificial intelligence–enabled analysis of public attitudes on Facebook and Twitter toward COVID-19 vaccines in the United Kingdom and the United States: Observational study, *J Med Internet Res*, **23(4)** (2021) e26627, https://doi.org/10.2196/26627.

6 Liu B, Sentiment analysis and subjectivity, in *Handbook of Natural Language Processing*, **2** (2010) 627–666, https://doi.org/10.1201/9781420085938-36.

7 Mäntylä M V, Graziotin D & Kuutila M, The evolution of sentiment analysis—A review of research topics, venues, and top cited papers, *Comput Sci Rev*, **27** (2018) 16–32, https://doi.org/10.1016/j.cosrev.2017.10.002.

8 Saxena A, Khanna A & Gupta D, Emotion recognition and detection methods: A comprehensive survey, *J Artif Intell Syst*, **2(1)** (2020) 53–79, https://doi.org/10.33969/ais.2020.21005.

9 Küçük D & Can F, Stance detection: A survey, *ACM Comput Surv*, **53(1)** (2020) 1–37, https://doi.org/10.1145/3369026.

10 Mohammad S M, Sobhani P & Kiritchenko S, Stance and sentiment in tweets, *ACM T Internet Technol*, **17(3)** (2017) 1–23, https://doi.org/10.1145/3003433.

11 Qazi U, Imran M & Ofli F, GeoCoV19: A dataset of hundreds of millions of multilingual COVID-19 tweets with location information, *SIGSPATIAL Special*, **12(1)** (2020) 6–15, https://doi.org/10.1145/3404820.3404823.

12 DeVerna M R, Pierri F, Truong B T, Bollenbacher J, Axelrod D, Loynes N, Torres-Lugo C, Yang K C, Menczer F & Bryden J, CoVaxxy: A collection of English-language Twitter posts about COVID-19 vaccines, *Int AAAI Conf Web Social Media*, **15** (2021) 992–999.

13 Pathuri S K, Anbazhagan N, Joshi G P & You J, Feature-based sentimental analysis on public attention towards COVID-19 using CUDA-SADBM classification model, *Sensors*, **22(1)** (2022) 80, https://doi.org/10.3390/s22010080.

14 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).

15 Pierri F, Tocchetti A, Corti L, Giovanni M D, Pavanetto S, Brambilla M & Ceri S, VaccinItaly: Monitoring Italian conversations around vaccines on Twitter and Facebook, *arXiv preprint arXiv:2101.03757*(2021).

16 Müller M & Salathé M, Addressing machine learning concept drift reveals declining vaccine sentiment during the COVID-19 pandemic, *arXiv preprint arXiv:2012.02197* (2020).

17 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).

18 Saleh SN, McDonald S A, Basit M A, Kumar S, Arasaratnam R J, Perl T M, Lehmann C U & and Medford R J, Public perception of COVID-19 vaccines through analysis of Twitter content and users, *medRxiv* (2021) 2021–04.

19 Stella M, Vitevitch M S & Botta F, Cognitive networks identify the content of English and Italian popular posts about COVID-19 vaccines: Anticipation, logistics, conspiracy and loss of trust, *arXiv preprint arXiv: 2103.15909* (2021).

20 Wani M A, Agarwal N & Bours P, Impact of unreliable content on social media users during COVID-19 and stance detection system, *Electronics*, **10(1)** (2020) 5, https://doi.org/10.3390/electronics10010005.

21 Kwok S W H, Vadde S K & Wang G, Tweet topics and sentiments relating to COVID-19 vaccination among Australian Twitter users: Machine learning analysis, *J Med Internet Res*, **23(5)** (2021) e26953, https://doi.org/10.2196/26953.

22 Lyu J C, Han E L & Luli G K, COVID-19 vaccine–related discussion on Twitter: topic modeling and sentiment analysis, *J Med Internet Res*, **23(6)** (2021) e24435, https://doi.org/10.2196/24435.

23 Boucher J C, Cornelson K, Benham J L, Fullerton M M, Tang T, Constantinescu C, Mourali M, Oxoby R J, Marshall D A, Hemmati H & Badami A, Analyzing social media to explore the attitudes and behaviors following the announcement of successful COVID-19 vaccine trials: Infodemiology study, *JMIR Infodemiol*, **1(1)** (2021) e28800, https://doi.org/10.2196/28800.

24 Niu Q, Liu J, Kato M, Shinohara Y, Matsumura N, Aoyama T & Nagai-Tanima M, Public opinion and sentiment before and at the beginning of COVID-19 vaccinations in Japan: Twitter analysis, *JMIR Infodemiol*, **2(1)** (2022) e32335, https://doi.org/10.2196/32335.

25 Benis A, Chatsubi A, Levner E & Ashkenazi S, Change in threads on Twitter regarding influenza, vaccines, and vaccination during the COVID-19 pandemic: Artificial intelligence–based infodemiology study, *JMIR Infodemiol*, **1(1)** (2021) e31983, https://doi.org/10.2196/31983.

26 Wang G & Kwok S W H, Using k-means clustering method with Doc2Vec to understand the Twitter users' opinions on COVID-19 vaccination, *IEEE EMBS Int Conf Biomed Health Informat* (IEEE) 2021, 1–4.

27 Jiang L C, Chu T H & Sun M, Characterization of vaccine tweets during the early stage of the COVID-19 outbreak in the United States: topic modeling analysis, *JMIR Infodemiol*, **1(1)** (2021) e25636, https://doi.org/10.2196/25636.

28 Karami A, Zhu M, Goldschmidt B, Boyajieff H R & Najafabadi M M, COVID-19 vaccine and social media in the US: Exploring emotions and discussions on Twitter, *Vaccines*, **9(10)** (2021) 1059, https://doi.org/10.3390/vaccines9101059.

29 Hu T, Wang S, Luo W, Zhang M, Huang X, Yan Y, Liu R, Ly K, Kacker V, She B & Li Z, Revealing public opinion towards COVID-19 vaccines with Twitter data in the United States: Spatiotemporal perspective, *J Med Internet Res*, **23(9)** (2021) e30854, https://doi.org/10.2196/30854.

30 Cascini F, Pantovic A, Al-Ajlouni Y A, Failla G, Puleo V, Melnyk A, Lontano A & Ricciardi W, Social media and attitudes towards a COVID-19 vaccination: A systematic review of the literature, *eClin Med*, **48** (2022) 101454, https://doi.org/10.1016/j.eclinm.2022.101454.

31 Küçük D & Arıcı N, Sentiment, stance, and emotion analysis on Twitter for COVID-19 vaccination: A survey, *Artif Intell Stud*, **5(1)** (2022) 14–22.

32 Qorib M, Oladunni T, Denis M, Ososanya E & Cotae P, COVID-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset, *Expert Syst Appl*, **212** (2023) 118715, https://doi.org/10.1016/j.eswa.2022.118715.

33 Sv P, Tandon J & Hinduja H, Indian citizen's perspective about side effects of COVID-19 vaccine–A machine learning study, *Diabetes Metab Syndr: Clin Rese Rev*, **15(4)** (2021) 102172, https://doi.org/10.1016/j.dsx.2021.06.009.

34 Huangfu L, Mo Y, Zhang P, Zeng D D & He S, COVID-19 vaccine tweets after vaccine rollout: Sentiment–based topic modeling, *J Med Internet Res*, **24(2)** (2022) e31726, https://doi.org/10.2196/31726.

35 Hutto C & Gilbert E, VADER: A parsimonious rule-based model for sentiment analysis of social media text, *Intl AAAI Conf Web Social Media*, **8(1)** (2014) 216–225.

36 Yin H, Song X, Yang S & Li J, Sentiment analysis and topic modeling for COVID-19 vaccine discussions, *World Wide Web*, **25(3)** (2022) 1067–1083, https://doi.org/10.1007/s11280-022-01029-y.

37 Küçük D & Arıcı N, Sentiment analysis and stance detection in Turkish tweets about COVID-19 vaccination, in *Handbook of Research on Opinion Mining and Text Analytics on Literary Works and Social Media* (IGI Global) 2022, 371–387.

38 Yu A W, Dohan D, Le Q, Luong T, Zhao R & Chen K, Norouzi M, Le Q V, Fast and accurate reading comprehension by combining self-attention and convolution, *Int Conf Learn Represent*, **2(1)** (2018).

39 Wei J & Zou K, EDA: Easy data augmentation techniques for boosting performance on text classification tasks, *arXiv preprint arXiv:1901.11196* (2019).

40 Xiang R, Chersoni E, Lu Q, Huang C R, Li W & Long Y, Lexical data augmentation for sentiment analysis, *J Assoc Inf Sci Technol*, **72(11)** (2021) 1432–1447, https://doi.org/10.1002/asi.24493.

41 Li Y & Caragea C, Target-aware data augmentation for stance detection, *Conf North Am Chapter Associat Comput Linguists: Human Language Technol* (Association for Computational Linguistics) 2021, 1850–1860.

42 Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P & Witten I H, The WEKA data mining software: An update. *ACM SIGKDD Explorations Newsletter*; **11(1)** (2009) 10–18, https://doi.org/10.1145/1656274.1656278.

43 Hochreiter S & Schmidhuber J, Long short-term memory, *Neural Comput*, (1997) **9(8)** 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735.

44 Lang S, Bravo-Marquez F, Beckham C, Hall M & Frank E, Weka Deeplearning4j: A deep learning package for Weka based on Deeplearning4j, *Knowl Based Syst*, **178** (2019) 48–50, https://doi.org/10.1016/j.knosys.2019.04.013.

45 Mohammad S, Kiritchenko S, Sobhani P, Zhu X & Cherry C, A dataset for detecting stance in tweets, in *Proc Tenth Int Conf Language Resour Evaluat* (European Language Resources Association) 2016, 3945–3952.

46 Priya S, Senthilkumar R & Jeyakumar S, Multi-label convolution neural network for personalized news recommendation based on social media mining, *J Sci Ind Res*, **81(7)** (2022) 785–97, https://doi.org/10.56042/jsir.v81i07.46261.

47 Küçük D, Arıcı N & Küçük D, Named entity recognition in Turkish: Approaches and issues, *Int Conf Appl Natural Language to Informat Syst* (Springer International Publishing) 2017, 176–181.