Leveraging Socio-contextual Information in BERT for Fake Health News Detection in Social Media

Rishabh Upadhyay
rishabh.upadhyay@unimib.it
University of Milano-Bicocca, Department of Informatics, Systems, and Communication, IKR3 Lab
Milan, Italy

Gabriella Pasi
gabriella.pasi@unimib.it
University of Milano-Bicocca, Department of Informatics, Systems, and Communication, IKR3 Lab
Milan, Italy

Marco Viviani*
marco.viviani@unimib.it
University of Milano-Bicocca, Department of Informatics, Systems, and Communication, IKR3 Lab
Milan, Italy

ABSTRACT

Fake news is a major challenge in social media, particularly in the health domain where it can lead to severe consequences for both individuals and society as a whole. To contribute to combating this problem, we present a novel solution for improving the accuracy of detecting fake health news, utilizing a fine-tuned BERT model that integrates both user- and content-related socio-contextual information. Specifically, this information is combined with the textual content itself to form a socio-contextual input sequence for the BERT model. By fine-tuning such a model with respect to the health misinformation detection task, the resulting classifier can accurately predict the category to which each piece of content belongs, i.e., either "real health news" or "fake health news". We validate our solution through a series of experiments conducted on distinct publicly available datasets constituted by health-related tweets. These results illustrate the superiority of the proposed solution compared to the standard BERT baseline model and other advanced models. Indeed, they show that the integration of socio-contextual information in the detection process positively contributes to increasing the overall accuracy of the fake health news detection task. The study also suggests, in a preliminary way, how such information could be used for the explainability of the model itself.

CCS CONCEPTS

• **Human-centered computing** → **Social media**; • **Computing methodologies** → **Natural language processing**; *Supervised learning by classification*; • **Applied computing** → **Consumer health**; • **Information systems** → *Content analysis and feature selection.*

KEYWORDS

Fake News, Health Misinformation, Transformers, BERT, Classification, Social Media, Social Context

*Corresponding author.

OASIS '23, September 04–08, 2023, Rome, Italy

ACM ISBN 979-8-4007-0225-9/23/09.
https://doi.org/10.1145/3599696.3612902

ACM Reference Format:
Rishabh Upadhyay, Gabriella Pasi, and Marco Viviani. 2023. Leveraging Socio-contextual Information in BERT for Fake Health News Detection in Social Media. In *3rd International Workshop on Open Challenges in Online Social Networks (OASIS '23), September 04–08, 2023, Rome, Italy.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3599696.3612902

1 INTRODUCTION

The COVID-19 pandemic, a significant global public health crisis that is considered the most substantial challenge since World War II, has resulted in over 75 million confirmed cases and 6.8 million deaths as of March 2023.[1] Additionally, it has led to the emergence of a real *infodemic*. The term was originally used by the *World Health Organization* (WHO) during the 2003 SARS outbreak [15], and has since been used to describe the spread of false information during other health crises. An infodemic can cause confusion, panic, and harm by promoting misleading or dangerous information that can undermine efforts to address and manage a crisis. The spread of fake news related to COVID-19 and other pandemics, either driven by individuals seeking economic or political gains [23, 41], or simply by people without sufficient health literacy [5, 39], has posed significant challenges for governments, global health organizations, and news outlets. Even without referring to pandemics, since these days social media platforms are a primary source of information for many people in different contexts including the health domain [35], the spread of false information, also exacerbated by the phenomenon of echo chambers [59], has become a critical issue both for the welfare of individuals and for that of society as a whole [12].

In this scenario, the detection and prevention of fake health news are essential to ensure that health information disseminated through social media is as much reliable and accurate as possible. Indeed, the use of Machine Learning tools to help laypeople discern the truthfulness of health information received from social media is increasingly necessary [7, 16, 25]. Over the past few years, health information shared on social media has been assessed by means of distinct solutions with respect to different but partly overlapping or related concepts, such as credibility [20, 49], completeness [48], relevance [54], readability [35], accuracy [24], veracity [64], etc. Many studies have aimed to create techniques to verify and detect fake health news on social media [11], and systematic literature reviews have provided useful insights into the challenges and opportunities in this field [3, 60, 62]. Many of these recent works have particularly exploited the potential of models based on *Bidirectional Encoder Representations from Transformers* (BERT) [30, 44, 47] and

[1] https://www.worldometers.info/coronavirus/

other neural architectures [28, 58]. However, they do not account for the use of full socio-contextual information by incorporating it in such models – an aspect that we believe is crucial for improving their effectiveness – and do not take into consideration in any way the problem of interpretability of the models themselves.

In this paper, therefore, we propose and study a solution to enhance the effectiveness of BERT in detecting fake health news on social media platforms, building on previous studies in the literature, but proposing an effective novelty. Indeed, our approach involves adjusting the input of the BERT model by including in it socio-contextual information related to both the user and their textual content posted on social media. Such an input sequence is subsequently fed into the BERT model, which is fine-tuned on the health misinformation detection task in order to leverage on such socio-contextual information to make the final prediction. The use of this socio-contextual information could be further used to understand its actual effectiveness with respect to the task of fake health news detection. The outcomes of our experiments, which are performed on publicly available health-related tweets, illustrate that our suggested approach surpasses the performance of the baseline BERT model and other advanced baselines, and achieves the best performance on fake health news datasets. Additionally, we have conducted a preliminary analysis to illustrate the possibility to provide interpretability of our model by using the *SHapley Additive exPlanations* (SHAP) game theoretic approach [36]. Our study reveals that the inclusion of socio-contextual information in BERT models offers further insights to differentiate between real and fake health news, and this could positively contribute to the dissemination of trustworthy and dependable health information on social media platforms.

In the following, Section 2 illustrates the main literature linguistic solutions that have addressed the problem of health misinformation on the Web and social media; Section 3 describes the proposed socio-contextual BERT model for fake health news detection; Section 4 describes the experimental evaluation, specifically the datasets used, some implementation details, the baselines used, and the results obtained; finally, Section 5 concludes the article and defines some possible future research directions.

2 RELATED WORK

The prevalence of fake news related to health has been increasing in recent years [18]. This type of fake news is often crafted with the intention to mislead the audience through the use of language that is sensational or opinionated [55, 62]. Indeed, researchers have identified certain linguistic features including the use of exaggerated language, the presentation of opinions as facts, and the use of emotionally charged language, which can help to identify such fake health news. For example, traditional *Machine Learning* (ML) *models* in association with *Bag-of-Words* (BoW) *features* (possibly weighed with TF-IDF), were employed by Aphinyanaphongs et al. [2] to detect Web pages that make unproven medical claims. Both *content-based* and *context-based features*, such as the length of the tweet, the tweet's sentiment, the number of hashtags, the number of followers of the tweet's author, the links present in the tweet, etc., were utilized to detect fake health news using *Decision Trees* (DTs) in [10]. Some *user features*, including the number of tweets a user has favored, the number of the user's tweets, and the user's description text, were used in [6, 27] to detect the trustworthiness of the users who wrote health-related news by applying *Random Forests* (RFs) and DTs. Defining *linguistic cues of deception*, i.e., sentence length, sentiment, emoticon usage, sentence complexity, and text-informality, RFs have been employed to detect fake health news using linguistic cues in [8]. *Logistic Regression* (LR) has also exhibited competitive performance in detecting fake health news in distinct proposals in association with both *linguistic* and *user-based features* [8, 9, 56]. *Stylistic features*, i.e., related to the syntax, text style, and grammatical elements of each article content and title, *complexity features*, i.e., based on deeper natural language processing computations to capture the overall intricacy of an article or title, and *psychological features*, i.e., based on well-studied word counts that are correlated with different psychological processes, and basic *sentiment analysis*, were used by Horne and Adali [26] to differentiate between fake news, satire news, and real news in the health domain. An enhanced evolutionary detection approach was proposed by Al-Ahmad et al. [1] to identify COVID-19 misinformation. The authors implemented three *wrapper feature selection* algorithms, i.e., *Particle Swarm Optimization* (PSO) [61], a *Genetic Algorithm* (GA), and the *Salp Swarm Algorithm* (SSA) [31]. These algorithms were utilized to reduce the number of symmetrical (redundant) features, such as BoW, *Term Frequency* (TF), and *Term Frequency-Inverse Document Frequency* (TF-IDF) features. In this way, the authors aimed to optimize the feature selection process and improve the accuracy of the detection model.

Recently, *neural network models* have attracted considerable interest due to advances in computational hardware and the availability of large amounts of data. These models have shown promising results in text representation and classification tasks, including misinformation or fake news detection, w.r.t. traditional ML models. In the study by Tashtoush et al. [57], deep neural networks such as *Long Short Term Memory* (LSTM) and *Bidirectional* LSTM (Bi-LSTM) networks, *Convolutional Neural Networks* (CNNs), and a hybrid solution composed of CNNs and LSTM networks, were employed to identify COVID-19-related fake news on social media platforms. Among the models, CNNs outperformed the others. Similarly, Shushkevich and Cardiff [53] compared classical Machine Learning algorithms, the GMDH-Shell tool,[2] and an LSTM network, for detecting fake news related to Coronavirus on small datasets. GMDH-Shell algorithms were found to outperform traditional Machine Learning algorithms. Pavlov and Mirceva [44] focused on detecting COVID-19-related fake news using pre-trained BERT[3] and RoBERTa[4] models. A multi-domain fake news detection model was proposed by Nan et al. [40] to address the challenge of domain shift, obtaining a benchmark of fake health news dataset for *Multi-Domain Fake News Detection* (MFND) with domain label annotated. Additionally, Mayank et al. [37] proposed DEAP-FAKED, a knowledge graph fake health news detection framework that combines NLP and GNN techniques to detect fake health news, achieving significant improvement in the F1 score on two publicly available datasets. Kumari et al. [32] proposed a self-ensemble SciBERT [4] model for detecting health misinformation in news, using a custom

[2]https://gmdhsoftware.com/
[3]https://huggingface.co/digitalepidemiologylab/covid-twitter-bert-v2
[4]https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment

dataset that combines existing FakeHealth dataset [14], and health articles from news fact-checking Website *Snopes.com*.

However, none of the above-mentioned neural network models leveraged the entire socio-contextual information related to health news, which is crucial in our opinion for improving the effectiveness of detecting health-related fake news on social media. Hence, in the next section, we introduce a new model, which incorporates socio-contextual features like user information and content-related engagement into the BERT model to leverage such information. Both to show how the proposed solution works and to evaluate it, we instantiate the model using textual content and related socio-contextual information from the Twitter platform.

3 A SOCIO-CONTEXTUAL BERT MODEL FOR FAKE HEALTH NEWS DETECTION

The proposed socio-contextual BERT model for fake health news detection is based on the modification of the *input sequence* of the BERT model in a way that takes into account the social context in which the text was produced. Specifically, we consider the social context of the tweet text and of its author.

Specifically, here are the steps we have followed to design the proposed model:

(1) Text preprocessing;
(2) Identification of the relevant socio-contextual information;
(3) Combination of the textual input and socio-contextual information;
(4) Fine-tuning of the BERT model on the health misinformation detection task on publicly available datasets;
(5) Preliminary interpretation of the predictions.

3.1 Text Preprocessing

The initial stage of text preprocessing is text cleaning, which entails eliminating *noise* and *irrelevant information* from tweet text. In our research, we eradicated punctuation and substituted email with `<EMAIL>`, URLs with `<URL>`, numbers with `<NUMBER>`, phone numbers with `<PHONE>`, currency symbols with `<CUR>`, and mentions (`@username`) from the tweet text. URLs and mentions are extraneous to the text's meaning and can distort the data.

Another crucial step in data preprocessing is managing *emojis*. Emojis are extensively used in social media platforms and can convey emotions and sentiments, making them valuable for fake news detection. However, handling emojis can be difficult as they are not part of the standard English vocabulary. In our research, we transformed emojis to their corresponding textual descriptions using the Python `emoji` library.[5] For instance, the "face with medical mask" emoji can indicate that the tweet relates to health, which can be easily managed by the model. A simple example of preprocessing text according to the rules outlined in this section is shown in Figure 1.

3.2 Socio-contextual Information Identification

To enhance the accuracy of identifying health fake news on social media, it is crucial to take into account not only the tweet's content but also other socio-contextual factors. Social engagements, such

[5]https://pypi.org/project/emoji/

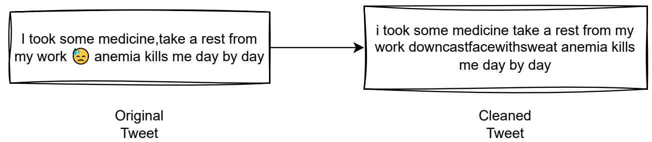

Figure 1: Example of preprocessing on textual content from tweets.

as the dissemination and interactions of news on social media, can offer valuable supplementary data to determine the truthfulness of news articles [22, 51]. To achieve this, we integrated the following contextual information associated with both the user and the tweet post.

3.2.1 *User Contextual Information.*

Although many social media users are real and act in a reliable way, there are some who are malicious or controlled by automated *social bots* that are managed by algorithms [21, 51]. As a result, gathering information about users' profiles and characteristics by utilizing user-based features can offer valuable information for detecting fake news. User-based features are those that represent the characteristics of users who interact with the news on social media, such as the number of followers/followees and the verified account status [9] that we considered in this work. The complete list of these features is detailed in Section 3.3.

3.2.2 *Tweet Contextual Information.*

Social media users frequently convey their emotions or views on fake news through posts, which may include skeptical opinions, sensational reactions, and more. As a result, it is sensible to extract post-based features to aid in the identification of potentially fake news through public reactions expressed in posts. In our investigation, we also obtained tweet-related contextual information, such as the number of retweets and likes. These statistics can indicate the tweet's popularity and potentially reveal suspicious activity or trends [34, 50, 52]. The complete list of these features is detailed in Section 3.3.

In Figure 2 we can see a simple example of how this socio-contextual information was extracted relative to users and tweets, before being fed into the BERT model suitably modified to accept this type of information.

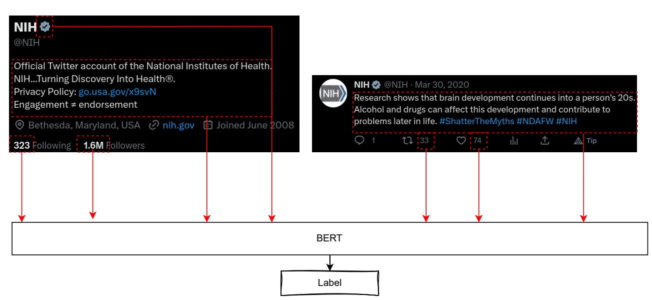

Figure 2: Extraction of socio-contextual information from tweets.

3.3 Socio-contextual Information Injection in the Input Sequence

After completing the data preprocessing and obtaining socio-contextual information related to both the user and the tweet, we adjusted the input of the BERT model to integrate this information. The typical input sequence for the BERT model consists of the following three components:

- The `[CLS]` token, which represents the start of the input and is used to obtain a fixed-size representation of the entire input sequence;
- The tokenized `Text` of the input sequence, which contains the actual text data. This text is typically split into words;
- The `[SEP]` token, which separates different sentences or segments within the input. This token can also be used to separate the tokenized text from other input features, such as contextual information.

Hence, the typical input is given as:

```
[CLS] Text [SEP]
```

To include both user-related and tweet-related features (i.e., related to user and tweet socio-contextual information), the new BERT input sequence proposed in this paper takes the following form:

```
[CLS] User Description [SEP] User Followers [SEP]
User Followees [SEP] Verified Status [SEP] Tweet Likes
[SEP] Number of Tweet Retweets [SEP] Tweet Text [SEP]
```

In the modified input sequence, `User Description` represents the description of the user account, `User Followers` represents the number of followers of a user, `User Followees` represents the number of their followees, `Verified Status` represents the verification status of the user account, `Tweet Likes` represents the number of likes on the tweet, `Number of Tweet Retweets` represents the number of retweets of the tweet, and `Tweet Text` represents the text of the tweet (eventually preprocessed). For example, if the text of the tweet was, "COVID-19 vaccines are now available in all pharmacies", with the following description related to the user, "I am a doctor with M.D. and work in a hospital", with 1,000 followers, 500 followers, and verified account status, and the information related to the tweet included 50 likes and 20 retweets, the modified input sequence would be as follows:

```
[CLS] i am a doctor with md and working
in a hospital [SEP] normalized(1000) [SEP]
normalized(500) [SEP] True [SEP] normalized(50) [SEP]
normalized(20) [SEP] covid19 vaccines are now
available in all pharmacies [SEP]
```

In this example, `normalized(*)` indicates the fact that numerical features are normalized using one of the normalization techniques analyzed in this paper and illustrated in Table 2. The modified input sequence is then fed into the BERT model, which learns to jointly represent the text content and socio-contextual information.

3.4 Fine-tuning the BERT model

The fine-tuning of the BERT model involves training it to predict tweet labels as real or fake news, using the modified input sequence with socio-contextual information. The objective function is the *cross-entropy loss*, with parameter updates during training done using the *Adam optimizer* at a learning rate of `2e-5`. The *batch size* is set at 8, while the *maximum sequence length* is 256. To avoid overfitting and ensure model robustness, a *five-fold cross-validation* technique is employed. The dataset is partitioned randomly into five equal subsets, with each subset used once for testing and the remaining four subsets for training. The model is trained five times, with each iteration using a different subset for testing.

3.5 Preliminary Prediction Interpretation

In addition to fine-tuning the BERT model, we leveraged SHAP (*SHapley Additive exPlanations*) values to gain insights into how the model arrived at its predictions [36]. By preliminarily analyzing the SHAP values, we were able to observe, for a subset of tweets, which features played the most significant role in the model's decision-making process. This allowed us to better understand how the model was interpreting the input data and what factors were driving its predictions. Overall, using SHAP values can be used to provide valuable information for interpreting and explaining the model's behavior with respect to socio-contextual information.

4 EXPERIMENTAL EVALUATION

This section is devoted to illustrating the experimental evaluations we carried out in this article. In particular, through these evaluations, we want to answer the following research questions:

- *R1. What is the level of pre-processing that makes the proposed BERT model acting on just textual information more effective for identifying fake health news?*
- *R2. What benefits does the proposed BERT model gain from considering both user and tweet socio-contextual information, instead of relying only on textual tweets? Also, what normalization technique is optimal for numerical features in our task?*
- *R3. To what extent can the proposed BERT model with socio-contextual information improve the detection of fake health news on social media platforms like Twitter, compared to baseline solutions?*
- *R4. Is it possible to interpret the proposed model with respect to the effectiveness that socio-contextual information has in enabling the detection of fake health news?*

However, before answering these questions (this will be done in Section 4.4), we intend to illustrate the publicly accessible datasets on which we evaluated our model, some hardware and implementation details, and the baselines we took as a reference to evaluate the effectiveness of the proposed solution with respect to certain evaluation metrics.

4.1 Datasets

Three datasets of public health fake news were used in our experiments. The first dataset, named CMU-MisCOV19 [38], includes 4,573 Twitter posts about COVID-19 categorized into 17 themes. In the dataset construction, tweets labeled as "True Treatment", "True Prevention", and "True Public Health Response" were considered as real news, whereas those labeled as "Fake Cure", "Fake Treatment", "False Fact or Prevention", and "False Public Health Response" were classified as fake news. Out of the total posts in this dataset, we analyzed 630 tweets still existing online, together with related socio-contextual information.

The second dataset is CoAID [13], a COVID-19 health misinformation dataset that consists of COVID-19 articles and COVID-19 claims. The COVID-19 articles contain reliable COVID-19 medical news and fact-checking articles, including both medical and non-medical concepts. Specifically, it includes 4,251 news and 296,000 related user engagements, ranging from December 1st, 2019 to September 1st, 2020. Since, over the years, most of the tweets were either removed or deleted from Twitter, we experimented with 1,632 real news tweets and 544 fake news tweets and related socio-contextual information.

Lastly, we employed the *FakeHealth* dataset [14], which comprises news content and expert-generated reviews linked to health and medical interventions from *HealthNewsReview.org*, a fact-checking organization for health-related news from mainstream US media. Specifically, it contains news content with rich features (e.g., text, image, tags), news reviews with detailed explanations (e.g., labels, explanations, news URL), social engagements, and a user-user social network. Finally, we had access to a total of 5,000 still available tweets and related socio-contextual information.

We took stringent measures to prevent information leakage between training and test splits. This was achieved by ensuring a strict partition between the training and testing sets. Importantly, we also ensured that there were no common authors between the training and test datasets to avoid bias and overfitting. This measure ensures that our model's performance is unbiased and its learning is not influenced by any repeated author-specific patterns, therefore maintaining the integrity of our results and enhancing the generalizability of our proposed model.

4.2 Hardware and Implementation Details

To perform experiments, we used a single machine with the following hardware characteristics:

- CPU: AMD Ryzen 7 5800h with Radeon graphics;
- GPU: GeForce RTX 3070 Mobile/Max-Q;
- RAM: 16 GB DDR4.

Our models were trained and tested using the *PyTorch* framework.[6] We employed the pre-trained BERT model, which we loaded from the *Hugging Face* `Transformers` library with weights pre-configured for three models those are BERT-base,[7] BioBERT,[8] and BERTweet-Covid,[9] listed in Table 3. As anticipated in Section 3.4, to fine-tune the BERT model we used the Adam optimizer with a learning rate of `2e-5` and a batch size of 8. The training process was limited to a maximum of 5 epochs.

4.3 Baselines and Evaluation Metrics

In addition to different configurations of the proposed model, which will be directly presented in Section 4.4, we have considered several baselines that are divided into three macro-families: simple Machine Learning models, Deep Neural Network models, and BERT models. They are denoted as follows:

- *ML-baselines*. They are constituted by two popular ML algorithms: *Support Vector Machines* (SVMs) and *Logistic Regression* (LR), acting on a simple TF-IDF representation of the tweet text. These baselines are widely used in the fake health news detection task [19, 43, 46, 53, 65];
- *DNN-baselines*. They are constituted by two popular word embedding techniques: *GloVe* and *Mittens*, with two different architectures, GRU and BiGRU as presented in [58] for fake health news detection. GloVe [45] is an unsupervised learning algorithm that can generate vector representations for words based on their co-occurrence statistics in the text corpus. Mittens [17] is an extension of the GloVe model that allows training with missing word pairs. Both GloVe and Mittens were used to obtain word embeddings for the tweet text. Then, we used these embeddings as input to the GRU and BiGRU models for classification;
- *BERT-baselines*. They are constituted by two previously proposed models for fake health news detection: FakeBERT and BERT+LSTM. FakeBERT [29] is a BERT model that uses a combination of BERT and CNNs to handle the textual content bidirectionally. BERT+LSTM [63] is a model that uses a pre-trained BERT model for text encoding and an LSTM network for classification.

The evaluation metrics used to comparatively compare our solution with these literature baselines are as follows: *accuracy*, F1 *score*, *recall*, and *precision*.

4.4 Results

In this section, through the presentation of the results obtained, we can answer the research questions introduced at the beginning of Section 4. *R1* and *R2* are preliminary research questions, which do not directly assess the effectiveness of the proposed model with respect to other baselines (this is entrusted to questions *R3* and *R4*). For this reason, experiments on the first two questions were conducted on a single dataset for detecting fake health news.

4.4.1 R1. What is the level of pre-processing that makes the proposed BERT model acting on just textual information more effective for identifying fake health news? In Table 1, the performance metrics including accuracy, F1 score, recall, and precision are compared on BERT-base for three distinct preprocessing techniques applied to the CMU-MisCOV19 dataset.

Table 1: Performance of a BERT model trained on different preprocessed data, i.e., raw tweets, tweet data without emoji, and tweet data with emoji conversion. The results (best are in bold) are compared using a statistical significance t-test.

	CMU-MisCOV19			
Data	Accuracy	F1	Recall	Precision
Tweet (no clean)	0.841	0.845	0.843	0.846
Tweet (cleaned) *without emojis*	0.870	0.870	0.870	0.861
Tweet (cleaned) *with emojis*	**0.875**	**0.875**	**0.878**	**0.870**

The first method involved using the raw tweet data, whereas the subsequent two methods involved preprocessing the data by

[6] https://pytorch.org/
[7] https://huggingface.co/bert-base-uncased
[8] https://huggingface.co/dmis-lab/biobert-v1.1
[9] https://huggingface.co/vinai/bertweet-covid19-base-uncased

eliminating irrelevant information and transforming emojis into textual form. The results illustrate that the model's performance was superior on the preprocessed data, as compared to the raw data which had not undergone any cleaning. Specifically, the accuracy of the model increased from 0.841 to 0.870 and 0.875 for the preprocessed tweet data (without emoji) and the preprocessed tweet data (with emoji), respectively. Across all four metrics, the preprocessed tweet data, both with and without emoji conversion, yielded better results than the raw tweet data. These results imply that data preprocessing is essential to enhance the accuracy of the model. Additionally, the model that employed emoji conversion demonstrated a slightly better performance than the model without emoji conversion, suggesting that converting emojis into text can offer supplementary insights that are valuable for fake news detection. Emojis are frequently utilized to express emotions or sentiments, which can be relevant in detecting fake news related to health issues.

4.4.2 R2. What benefits does the proposed BERT model gain from considering both user and tweet socio-contextual information, instead of relying only on textual tweets? Also, what normalization technique is optimal for numerical features in our task?

In Table 2, the effectiveness of the proposed BERT model (BERT-base) in identifying fake health news on Twitter has been evaluated using cleaned textual tweets (with emojis) only and in combination with different socio-contextual information. This contextual information comprises of user description, all user information such as followers, friends, verified status, and user description, as well as all tweet information including the number of retweets and likes. As there are no set limits for the number of retweets, likes, followers, and friends, we had to utilize various normalization methods, such as *z-score* and *min-max* normalization [42], to account for this variability.

Table 2: Performance of a BERT model using different combinations of socio-contextual information (SCI) and normalization (Norm.) of numerical data in association with cleaned textual tweets (CT). The results (best are in bold) are compared using a statistical significance t-test.

		CMU-MisCOV19			
CT + [SCI]	Norm.	Accuracy	F1	Recall	Precision
No SCI		0.875	0.875	0.878	0.870
User Descr.		0.891	0.885	0.890	0.880
All User Info	*z-score*	0.909	0.921	0.933	0.910
	min-max	0.891	0.896	0.894	0.898
All Tweet Info	*z-score*	0.906	0.911	0.919	0.903
	min-max	0.892	0.888	0.887	0.890
All User Info +	*z-score*	**0.923**	**0.923**	**0.946**	**0.919**
All Tweet Info	*min-max*	0.903	0.891	0.908	0.875

The results indicate that incorporating socio-contextual information enhances the model's performance. The most optimal performance was observed when all user and tweet information was combined, implying that the amalgamation of both types of information provides the most comprehensive and informative socio-contextual information for the model to make precise decisions.

More in detail, utilizing solely the user's description or all user information also results in improved performance when compared to solely using cleaned textual tweets. This highlights the usefulness of user information such as the user's description, followers, friends, and verified status in providing valuable contextual information to the model. Conversely, employing all tweet information alongside the tweet alone did not perform as well as using user information or a combination of both user and tweet information. This implies that tweet information alone may not be sufficient in providing enough socio-contextual information to the model.

Regarding the best normalization method, the results demonstrate that utilizing *z-score* normalization generally produces better performance than *min-max* normalization.

4.4.3 R3. To what extent can the proposed BERT model with socio-contextual information improve the detection of fake health news on social media platforms like Twitter, compared to baseline solutions?

Table 3 presents a comparison of the performance of baseline models and different configurations of our model for detecting fake health news. The evaluation is conducted on the three datasets detailed in Section 4.1. In particular, model results are categorized into four sections: (1) baseline results comparing Support Vector Machines (SVMs) and Logistic Regression (LR) models, (2) DNN-baseline results on DNN-based models employing both textual and socio-contextual information, (3) BERT-baseline results, and (4) results associated with variations of the proposed model with BERT-base, BioBERT, and BERTweet-Covid. In all four cases, the clean textual content (with emojis) of the tweets is used, together with the z-score method for the normalization of socio-contextual information.

The results indicate that the proposed BERT model taking into account socio-contextual information in addition to the text provides the most outstanding performance with respect to all baselines. In particular, all the tested model configurations utilizing BERT-base, BioBERT, and BERTweet-Covid exhibit the highest accuracy, F1 score, recall, and precision for all datasets.

4.4.4 R4. Is it possible to interpret the proposed model with respect to the effectiveness that socio-contextual information has in enabling the detection of fake health news?

To answer this question, as anticipated in Section 3.5, we used SHAP values to preliminarily assess the interpretability of our model. In the table illustrated in Figure 3, we present just a couple of examples of real and fake health news input sequences with their SHAP values. The red and blue colors in the table represent the positive and negative contributions of the features to the model's prediction, respectively. As can be observed, in both the rows of the table input sequence, the user-related information feature has a positive contribution (in red) for making real or fake health news predictions, as already uninterpretably revealed by the results shown in Table 2. due to space constraints, we have provided only an example image of such an interpretability mechanism applied to input sequences. Additional data can be found at the link provided in the Code and Data Availability Section.

5 CONCLUSIONS AND FUTURE WORK

Our research aimed to develop an innovative approach for identifying health-related fake news on social media platforms. To accomplish this, we integrated socio-contextual information about

Table 3: Evaluation of the effectiveness of the model w.r.t. to baselines. CT: cleaned textual tweets. SCI: socio-contextual information (All User Info + All Tweet Info). Results are referred to CMU-MisCOV19, CoAID, and FakeHealth. The results are compared using a statistical significance *t*-test.

			CMU-MisCOV19			
Results Category	Model	Data	Accuracy	F1	Recall	Precision
ML-baselines	SVMs	CT	0.743	0.703	0.711	0.694
	LR	CT	0.722	0.711	0.721	0.702
DNN-baselines	Glove_GRU	CT + SCI	0.800	0.801	0.801	0.801
	Glove_BiGRU	CT + SCI	0.804	0.807	0.806	0.808
	Mittens_GRU	CT + SCI	0.751	0.751	0.751	0.751
	Mittens_BiGRU	CT + SCI	0.763	0.765	0.765	0.766
BERT-baselines	FakeBERT	CT	0.885	0.879	0.878	0.871
	BERT+LSTM	CT	0.896	0.898	0.896	0.899
Proposed Model	BERT-base	CT + SCI	0.919	0.926	0.939	0.913
	BioBERT	CT + SCI	0.923	0.923	0.945	0.919
	BERTweet-Covid	CT + SCI	0.955	0.955	0.953	0.957
			CoAID			
Results Category	Model	Data	Accuracy	F1	Recall	Precision
ML-baselines	SVMs	CT	0.712	0.792	0.795	0.788
	LR	CT	0.700	0.798	0.801	0.795
DNN-baselines	Glove_GRU	CT + SCI	0.866	0.864	0.861	0.867
	Glove_BiGRU	CT + SCI	0.880	0.885	0.881	0.889
	Mittens_GRU	CT + SCI	0.847	0.856	0.856	0.856
	Mittens_BiGRU	CT + SCI	0.876	0.871	0.871	0.871
BERT-baselines	FakeBERT	CT	0.899	0.899	0.904	0.894
	BERT+LSTM	CT	0.913	0.909	0.912	0.906
Proposed Model	BERT-base	CT + SCI	0.943	0.939	0.934	0.945
	BioBERT	CT + SCI	0.975	0.961	0.966	0.956
	BERTweet-Covid	CT + SCI	0.985	0.979	0.978	0.980
			FakeHealth			
Results Category	Model	Data	Accuracy	F1	Recall	Precision
ML-baselines	SVMs	CT	0.434	0.619	0.767	0.500
	LR	CT	0.456	0.628	0.786	0.502
DNN-baselines	Glove_GRU	CT + SCI	0.545	0.701	0.867	0.567
	Glove_BiGRU	CT + SCI	0.550	0.714	0.883	0.577
	Mittens_GRU	CT + SCI	0.523	0.667	0.833	0.534
	Mittens_BiGRU	CT + SCI	0.534	0.686	0.851	0.554
BERT-baselines	FakeBERT	CT	0.604	0.739	0.894	0.612
	BERT+LSTM	CT	0.623	0.754	0.901	0.630
Proposed Model	BERT-base	CT + SCI	0.654	0.779	0.923	0.657
	BioBERT	CT + SCI	0.723	0.816	0.954	0.698
	BERTweet-Covid	CT + SCI	0.798	0.843	0.983	0.723

users and their tweets into the input sequence of a fine-tuned BERT model. The results of our experiments demonstrate that incorporating such information into the Transformer-based model significantly improves its ability to detect health-related fake news compared to existing state-of-the-art models. Notably, user-related features like the number of followers, followees, and verified status proved particularly valuable in identifying fake news. Furthermore, tweet-related features such as the number of retweets and likes

Input Sequence	Lab
social media marketing social media management and educating business owners on how to use twitter facebook and youtube for business mom and wife[SEP] 0.0557 [SEP] 0.011 [SEP] False [SEP] 0.134 [SEP] 0.1285 [SEP] study suggests osteoporosis drug might treat loss of bone in jaw saturday oct <number> healthday news people su <url>	1
most popular and emailed news managed by <url> [SEP] -0.0685 [SEP] -0.1143 [SEP] False [SEP] -0.134 [SEP] -0.1285 [SEP] study oks light drinking during pregnancy too good to be true <url>	0

Figure 3: Using SHAP values to interpret the model with respect to the contribution of socio-contextual information in the detection of fake health news.

also played a crucial role in fake news detection. Our preliminary analysis of SHAP values further suggests the potential importance of considering these socio-contextual features.

One promising avenue for future research involves investigating the effectiveness of various pre-training strategies, such as domain-specific pre-training or multi-task learning, to enhance the model's performance in detecting fake health news within specific contexts. Furthermore, exploring the adaptability and applicability of the proposed model to diverse languages and cultures could prove beneficial, considering the varying socio-contextual factors that influence the dissemination and perception of fake health news across regions and communities. Lastly, conducting a comprehensive and structured analysis of the model's interpretability would be worthwhile, aiming to gain deeper insights into its decision-making process and offer users more transparent information regarding its predictions.

CODE AND DATA AVAILABILITY

The datasets used within this work are publicly accessible and have been indicated in Section 4.1. The code for the reproducibility of the experiments is made available at the following address: https://github.com/ikr3-lab/socioBERTfake/.

USE OF INCLUSIVE LANGUAGE

For the purpose of inclusive writing, we have used the English form of the *singular they* [33] in this article.

ACKNOWLEDGMENTS

This work is supported by the EU Horizon 2020 ITN/ETN DoSSIER Project on Domain-Specific Systems for Information Extraction and Retrieval (H2020-EU.1.3.1., ID: 860721), https://dossier-project.eu/.

REFERENCES

[1] Bilal Al-Ahmad, Ala'M Al-Zoubi, Ruba Abu Khurma, and Ibrahim Aljarah. 2021. An evolutionary fake news detection method for COVID-19 pandemic information. *Symmetry* 13, 6 (2021), 1091.

[2] Yindalon Aphinyanaphongs, Constantin Aliferis, et al. 2007. Text categorization models for identifying unproven cancer treatments on the web. *Studies in Health Technology and Informatics* 129(Pt 2) (2007), 968–972.

[3] Manuel Au-Yong-Oliveira, Carlota P. A. Carlos, Hugo Pintor, João Caires, and Julia Zanoni. 2019. Fake News and Social Networks: How Users Interact with Fake Content. In *New Knowledge in Information Systems and Technologies*. Springer International Publishing, Cham, 195–205.

[4] Iz Beltagy, Kyle Lo, and Arman Cohan. 2019. SciBERT: A Pretrained Language Model for Scientific Text. arXiv:1903.10676 [cs.CL]

[5] Salman Bin Naeem and Maged N Kamel Boulos. 2021. COVID-19 misinformation online and health literacy: a brief overview. *International Journal of Environmental Research and Public Health* 18, 15 (2021), 8091.

[6] Todd Bodnar, Conrad Tucker, Kenneth Hopkinson, and Sven G. Bilén. 2014. Increasing the veracity of event detection on social media networks through user trust modeling. In *2014 IEEE International Conference on Big Data (Big Data)*. IEEE, 636–643. https://doi.org/10.1109/BigData.2014.7004286

[7] Alessandro Bondielli and Francesco Marcelloni. 2019. A survey on fake news and rumour detection techniques. *Information Sciences* 497 (2019), 38–55.

[8] Erica J. Briscoe, D. Scott Appling, and Heather Hayes. 2014. Cues to Deception in Social Media Communications. In *2014 47th Hawaii International Conference on System Sciences*. IEEE, 1435–1443. https://doi.org/10.1109/HICSS.2014.186

[9] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web*. ACM, 675–684.

[10] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2013. Predicting information credibility in time-sensitive social media. *Internet Research* 23, 5 (2013), 560–588.

[11] Canyu Chen, Haoran Wang, Matthew Shapiro, Yunyu Xiao, Fei Wang, and Kai Shu. 2022. Combating Health Misinformation in Social Media: Characterization, Detection, Intervention, and Open Issues. arXiv:2211.05289 [cs.SI]

[12] Liang Chen, Xiaohui Wang, and Tai-Quan Peng. 2018. Nature and diffusion of gynecologic cancer–related misinformation on social media: analysis of tweets. *Journal of Medical Internet Research* 20, 10 (2018), e11515.

[13] Limeng Cui and Dongwon Lee. 2020. CoAID: COVID-19 Healthcare Misinformation Dataset. arXiv:2006.00885 [cs.SI]

[14] Enyan Dai, Yiwei Sun, and Suhang Wang. 2020. Ginger Cannot Cure Cancer: Battling Fake Health News with a Comprehensive Data Repository. *Proceedings of the International AAAI Conference on Web and Social Media* 14, 1 (May 2020), 853–862. https://doi.org/10.1609/icwsm.v14i1.7350

[15] Rakesh Datta, AK Yadav, Anubhav Singh, Karuna Datta, and Ankur Bansal. 2020. The infodemics of COVID-19 amongst healthcare professionals in India. *Medical Journal Armed Forces India* 76, 3 (2020), 276–283.

[16] João Victor de Souza, Jorão Gomes Jr, Fernando Marques de Souza Filho, Alessandreia Marta de Oliveira Julio, and Jairo Francisco de Souza. 2020. A systematic mapping on automatic classification of fake news in social media. *Social Network Analysis and Mining* 10 (2020), 1–21.

[17] Nicholas Dingwall and Christopher Potts. 2018. Mittens: An Extension of GloVe for Learning Domain-Specialized Representations. arXiv:1803.09901 [cs.CL]

[18] Israel Junior Borges Do Nascimento, Ana Beatriz Pizarro, Jussara M Almeida, Natasha Azzopardi-Muscat, Marcos André Gonçalves, Maria Björklund, and David Novillo-Ortiz. 2022. Infodemics and health misinformation: a systematic review of reviews. *Bulletin of the World Health Organization* 100, 9 (2022), 544.

[19] Mohamed K. Elhadad, Kin Fun Li, and Fayez Gebali. 2020. Detecting Misleading Information on COVID-19. *IEEE Access* 8 (2020), 165201–165215. https://doi.org/10.1109/ACCESS.2020.3022867

[20] Gunther Eysenbach. 2008. *Credibility of health information and digital media: New perspectives and implications for youth.* MacArthur Foundation Digital Media and Learning Initiative.

[21] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2016. The rise of social bots. *Commun. ACM* 59, 7 (2016), 96–104.

[22] Matthew Gentzkow, Jesse M Shapiro, and Daniel F Stone. 2015. Media bias in the marketplace: Theory. In *Handbook of media economics*. Vol. 1. Elsevier, 623–645.

[23] Kacper Gradon. 2020. Crime in the time of the plague: fake news pandemic and the challenges to law-enforcement and intelligence community. *Society Register* 4 (2020), 133–148.

[24] M Nick Hajli, Julian Sims, Mauricio Featherman, and Peter ED Love. 2015. Credibility of information in online communities. *Journal of Strategic Marketing* 23, 3 (2015), 238–253.

[25] Ebtihal A Hassan and Farid Meziane. 2019. A survey on automatic fake news identification techniques for online and socially produced data. In *2019 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE)*. IEEE, 1–6.

[26] Benjamin Horne and Sibel Adali. 2017. This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In *Proceedings of the international AAAI conference on web and social media*, Vol. 11. AAAI, 759–766.

[27] Jun Ito, Jing Song, Hiroyuki Toda, Yoshimasa Koike, and Satoshi Oyama. 2015. Assessment of tweet credibility with LDA features. In *Proceedings of the 24th international conference on world wide web*. ACM, 953–958.

[28] Rohit Kumar Kaliyar, Anurag Goswami, and Pratik Narang. 2021. EchoFakeD: improving fake news detection in social media with an efficient deep neural network. *Neural Computing and Applications* 33 (2021), 8597–8613.

[29] Rohit Kumar Kaliyar, Anurag Goswami, and Pratik Narang. 2021. FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. *Multimedia Tools and Applications* 80, 8 (2021), 11765–11788.

[30] Ashfia Jannat Keya, Md Anwar Hussen Wadud, MF Mridha, Mohammed Alatiyyah, and Md Abdul Hamid. 2022. AugFake-BERT: Handling Imbalance through Augmentation of Fake News Using BERT to Enhance the Performance of Fake News Classification. *Applied Sciences* 12, 17 (2022), 8398.

[31] Ruba Abu Khurma, Khair Eddin Sabri, Pedro A Castillo, and Ibrahim Aljarah. 2021. Salp swarm optimization search based feature selection for enhanced phishing websites detection. In *Applications of Evolutionary Computation: 24th International Conference, EvoApplications 2021, Held as Part of EvoStar 2021, Virtual Event, April 7–9, 2021, Proceedings 24*. Springer, 146–161.

[32] Santoshi Kumari, Harshitha K Reddy, Chandan S Kulkarni, and Vanukuri Gowthami. 2021. Debunking health fake news with domain specific pre-trained model. *Global Transitions Proceedings* 2, 2 (2021), 267–272.

[33] Darren K LaScotte. 2016. Singular they: An empirical study of generic pronoun use. *American Speech* 91, 1 (2016), 62–80.

[34] David MJ Lazer, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, et al. 2018. The science of fake news. *Science* 359, 6380 (2018), 1094–1096.

[35] DM Lopez, Bernd Blobel, and C Gonzalez. 2016. Information quality in healthcare social media–an architectural approach. *Health and Technology* 6 (2016), 17–25.

[36] Scott M Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.). Curran Associates, Inc., 4765–4774. http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

[37] Mohit Mayank, Shakshi Sharma, and Rajesh Sharma. 2022. DEAP-FAKED: Knowledge Graph-based Approach for Fake News Detection. arXiv:2107.10648 [cs.CL]

[38] Shahan Ali Memon and Kathleen M. Carley. 2020. Characterizing COVID-19 Misinformation Communities Using a Novel Twitter Dataset. arXiv:2008.00791 [cs.SI]

[39] Andrea Moscadelli, Giuseppe Albora, Massimiliano Alberto Biamonte, Duccio Giorgetti, Michele Innocenzio, Sonia Paoli, Chiara Lorini, Paolo Bonanni, and Guglielmo Bonaccorsi. 2020. Fake news and Covid-19 in Italy: results of a quantitative observational study. *International Journal of Environmental Research and Public Health* 17, 16 (2020), 5850.

[40] Qiong Nan, Juan Cao, Yongchun Zhu, Yanyan Wang, and Jintao Li. 2021. MDFEND: Multi-Domain Fake News Detection. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management* (Virtual Event, Queensland, Australia) *(CIKM '21)*. Association for Computing Machinery, New York, NY, USA, 3343–3347. https://doi.org/10.1145/3459637.3482139

[41] Katherine Ognyanova, David Lazer, Ronald E Robertson, and Christo Wilson. 2020. Misinformation in action: Fake news exposure is linked to lower trust in media, higher trust in government when your side is in power. *Harvard Kennedy School Misinformation Review* 1 (2020), 19 pages.

[42] S. Gopal Krishna Patro and Kishore Kumar Sahu. 2015. Normalization: A Preprocessing Stage. arXiv:1503.06462 [cs.OH]

[43] Parth Patwa, Shivam Sharma, Srinivas Pykl, Vineeth Guptha, Gitanjali Kumari, Md Shad Akhtar, Asif Ekbal, Amitava Das, and Tanmoy Chakraborty. 2021. Fighting an infodemic: Covid-19 fake news dataset. In *Combating Online Hostile Posts in Regional Languages during Emergency Situation: First International Workshop, CONSTRAINT 2021, Collocated with AAAI 2021, Virtual Event, February 8, 2021, Revised Selected Papers 1*. Springer, 21–29.

[44] Tashko Pavlov and Georgina Mirceva. 2022. COVID-19 Fake News Detection by Using BERT and RoBERTa models. In *2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO)*. IEEE, 312–316.

[45] Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. GloVe: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Doha, Qatar, 1532–1543. https://doi.org/10.3115/v1/D14-1162

[46] Karishnu Poddar, KS Umadevi, et al. 2019. Comparison of various machine learning models for accurate detection of fake news. In *2019 Innovations in Power and Advanced Computing Technologies (i-PACT)*, Vol. 1. IEEE, 1–5.

[47] Nishant Rai, Deepika Kumar, Naman Kaushik, Chandan Raj, and Ahad Ali. 2022. Fake News Classification using transformer based enhanced LSTM and BERT. *International Journal of Cognitive Computing in Engineering* 3 (2022), 98–105.

[48] David Robins, Jason Holmes, and Mary Stansbury. 2010. Consumer health information on the Web: The relationship of visual design and perceptions of credibility. *Journal of the American Society for Information Science and Technology* 61, 1 (2010), 13–29.

[49] Charles C Self. 2014. Credibility. In *An integrated approach to communication theory and research*. Routledge, 449–470.

[50] Tracy Jia Shen, Robert Cowell, Aditi Gupta, Thai Le, Amulya Yadav, and Dongwon Lee. 2019. How gullible are you? Predicting susceptibility to fake news. In *Proceedings of the 10th ACM Conference on Web Science*. ACM, 287–288.

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

[52] Kai Shu, Suhang Wang, and Huan Liu. 2018. Understanding user profiles on social media for fake news detection. In *2018 IEEE conference on multimedia information processing and retrieval (MIPR)*. IEEE, 430–435.

[53] Elena Shushkevich and John Cardiff. 2021. Detecting fake news about Covid-19 on small datasets with machine learning algorithms. In *2021 30th Conference of Open Innovations Association FRUCT*. IEEE, 253–258.

[54] Shijie Song, Yan Zhang, and Bei Yu. 2021. Interventions to support consumer evaluation of online health information credibility: A scoping review. *International Journal of Medical Informatics* 145 (2021), 104321.

[55] Victor Suarez-Lledo and Javier Alvarez-Galvez. 2021. Prevalence of health misinformation on social media: systematic review. *Journal of Medical Internet Research* 23, 1 (2021), e17187.

[56] Eugenio Tacchini, Gabriele Ballarin, Marco L. Della Vedova, Stefano Moret, and Luca de Alfaro. 2017. Some Like it Hoax: Automated Fake News Detection in Social Networks. arXiv:1704.07506 [cs.LG]

[57] Yahya Tashtoush, Balqis Alrababah, Omar Darwish, Majdi Maabreh, and Nasser Alsaedi. 2022. A deep learning framework for detection of COVID-19 fake news on social media platforms. *Data* 7, 5 (2022), 65.

[58] Ciprian-Octavian Truică, Elena-Simona Apostol, and Panagiotis Karras. 2023. DANES: Deep Neural Network Ensemble Architecture for Social and Textual Context-aware Fake News Detection. arXiv:2302.01756 [cs.AI]

[59] Giacomo Villa, Gabriella Pasi, and Marco Viviani. 2021. Echo chamber detection and analysis: a topology-and content-based approach in the COVID-19 scenario. *Social Network Analysis and Mining* 11, 1 (2021), 78.

[60] Marco Viviani and Gabriella Pasi. 2017. Credibility in social media: opinions, news, and health information—a survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 7, 5 (2017), e1209.

[61] Feng Wang, Heng Zhang, and Aimin Zhou. 2021. A particle swarm optimization algorithm for mixed-variable optimization problems. *Swarm and Evolutionary Computation* 60 (2021), 100808.

[62] Yuxi Wang, Martin McKee, Aleksandra Torbica, and David Stuckler. 2019. Systematic literature review on the spread of health-related misinformation on social media. *Social Science & Medicine* 240 (2019), 112552.

[63] Apurva Wani, Isha Joshi, Snehal Khandve, Vedangi Wagh, and Raviraj Joshi. 2021. Evaluating deep learning approaches for covid19 fake news detection. In *Combating Online Hostile Posts in Regional Languages during Emergency Situation: First International Workshop, CONSTRAINT 2021, Collocated with AAAI 2021, Virtual Event, February 8, 2021, Revised Selected Papers 1*. Springer, 153–163.

[64] Elissa R Weitzman, Emily Cole, Liljana Kaci, and Kenneth D Mandl. 2011. Social but safe? Quality and safety of diabetes-related online social networks. *Journal of the American Medical Informatics Association* 18, 3 (2011), 292–297.

[65] Kasra Majbouri Yazdi, Adel Majbouri Yazdi, Saeid Khodayi, Jingyu Hou, Wanlei Zhou, and Saeed Saedy. 2020. Improving fake news detection using k-means and support vector machine approaches. *International Journal of Electronics and Communication Engineering* 14, 2 (2020), 38–42.