information

Article

Automated Classification of Fake News Spreaders to Break the Misinformation Chain

Simone Leonardi [1,*], Giuseppe Rizzo [2] and Maurizio Morisio [1]

[1] Politecnico di Torino DAUIN, Corso Duca degli Abruzzi, 24, 10129 Turin, Italy; maurizio.morisio@polito.it
[2] LINKS Foundation, Via Pier Carlo Boggio, 61, 10138 Turin, Italy; giuseppe.rizzo@linksfoundation.com
* Correspondence: simone.leonardi@polito.it; Tel.: +39-011-090-7087

Abstract: In social media, users are spreading misinformation easily and without fact checking. In principle, they do not have a malicious intent, but their sharing leads to a socially dangerous diffusion mechanism. The motivations behind this behavior have been linked to a wide variety of social and personal outcomes, but these users are not easily identified. The existing solutions show how the analysis of linguistic signals in social media posts combined with the exploration of network topologies are effective in this field. These applications have some limitations such as focusing solely on the fake news shared and not understanding the typology of the user spreading them. In this paper, we propose a computational approach to extract features from the social media posts of these users to recognize who is a fake news spreader for a given topic. Thanks to the CoAID dataset, we start the analysis with 300 K users engaged on an online micro-blogging platform; then, we enriched the dataset by extending it to a collection of more than 1 M share actions and their associated posts on the platform. The proposed approach processes a batch of Twitter posts authored by users of the CoAID dataset and turns them into a high-dimensional matrix of features, which are then exploited by a deep neural network architecture based on transformers to perform user classification. We prove the effectiveness of our work by comparing the precision, recall, and f_1 score of our model with different configurations and with a baseline classifier. We obtained an f_1 score of 0.8076, obtaining an improvement from the state-of-the-art by 4%.

Keywords: misinformation; social media; nlp; deep learning; sentence embeddings; natural language processing; multilingual embeddings; fake news; fact checking; user classification

Citation: Leonardi, S.; Rizzo, G.; Morisio, M. Automated Classification of Fake News Spreaders to Break the Misinformation Chain. *Information* **2021**, *12*, 248. https://doi.org/10.3390/info12060248

Academic Editors: Carlos A. Iglesias and J. Fernando Sánchez-Rada

Received: 7 May 2021
Accepted: 11 June 2021
Published: 15 June 2021

1. Introduction

Since the World Health Organization (WHO) declared COVID-19 a pandemic on 11 March 2020, social media platforms and traditional media have been flooded by information about the virus and behaviors to be followed to avoid its spread. At the same time, uncertainty and ambiguity regarding information about COVID-19 brought people to respond with non-adaptive coping strategies. In [1], Ha et al. stated that messaging to the public requires not only status reports and behavioral guidelines but also a component of positive information that can reduce anxiety. When this strategy does not work, people tend to react by generating harmful scenarios such as fake news production to protect themselves by trying to minimize the perceived danger. According to these findings, the COVID-19 pandemic increased the diffusion of wrong and misleading information on social media. In 2020, there was an exponential growth of cases in which a person received forged news and spread it rapidly on social media platforms without verifying its reliability. As reported by Cinelli et al. [2] and by the World Health Organization, https://www.who.int/news-room/spotlight/let-s-flatten-the-infodemic-curve (accessed on 15 March 2021), societies worldwide faced a parallel pandemic of fake information that required effective countermeasures to reduce the human effort needed to detect misinformation and to slow fake news diffusion. As reported in the survey by Oshikawa et al. [3],

language models have been widely employed to tackle this problem. In this context, Islam et al. explored multiple deep learning architectures to detect fake news [4]. In parallel, in [5], Jiang et al. investigated linguistic signals to find emotional markers in text, and they discovered a different social media interaction by the user when the user is encouraged to read fact-checked articles. In a similar scenario, Glenski et al. monitored different types of reactions to misinformation such as answer, appreciation, elaboration, and question [6]. Understanding the stance of a user about the content they share is a fundamental passage to effectively classify the user as a supporter or a detractor. All of these works reveal how fake news diffusion mechanisms are linked to the characteristics of the user sharing them. Actually, the definitions of fake news spreaders and checkers are a consequence of the fake news phenomenon on social media. Fake news spreaders are social media users supporting fake news and sharing misinformation. On the other hand, real news checkers are social media users sharing real news and supporting them. We describe the checked fake news as content declared false by fact-checking agencies after a human revision process.

The cited research projects report a growing need for automatic solutions to support fact checking agencies in their monitoring actions as well as to stimulate awareness of citizens in verifying content before sharing. In fact, when a user spreads fake information, they reinforce the trust of the community in the content, exponentially extending its reach, as explained by Vosoughi et al. [7]. When the information goes viral, the authorities must spend huge efforts to demonstrate its untruthfulness. The existing contributions to this research field show the effectiveness of language model adoptions combined with social media interaction analysis to detect misinformation. Even if these solutions have a big impact on society, we argue that they do not extensively analyze misinformation from the point of view of the user sharing it. An analysis of the user's posts and behavior on social media platforms fills this gap, expressing their perspective explicitly. The solutions proposed do not explore the encoding of the user's timeline into sentence embeddings to classify their tendency to share misinformation. In addition, they do not compare natural language processing approaches with respect to the machine learning models by exploiting social media graphs features.

Thus, we conducted our study by answering the following research questions:

RQ1 Is sentence encoding based on transformers and deep learning effective in classifying spreaders of fake news in the context of COVID-19 news?

RQ2 Which gold standard can be used for classifying spreaders of fake news in the context of COVID-19 news?

In this paper, we present the FNSC (Fake News Spreader Classifier), a stacked and transformer-based neural network that combines the transformer [8] capabilities of computing sentence embeddings with our deep learning model to classify users sharing fake news about COVID-19. This model transforms batches of tweets into sentence embeddings and processes them to classify users in a supervised approach. Starting from the dataset produced by Limeng and Dongwon [9], we collected tweet authors and their timelines to extensively inspect what they shared about COVID-19 and if they support the news they shared through a stance detection model. We show that our model has state-of-the-art results with the linguistic features. We also checked our model results using social media metrics alone, obtaining lower scores than the linguistic ones. The code we built is available in a publicly accessible repository: https://github.com/D2KLab/stopfaker (accessed on 7 May 2021). The CoAID dataset is also available in a publicly accessible repository: https://github.com/cuilimeng/CoAID (accessed on 16 January 2021). We shared the extended version of CoAID on Figshare, https://doi.org/10.6084/m9.figshare.14392859 (accessed on 7 May 2021).

The remainder of this paper is structured as follows. In Section 2, we illustrate how various studies approached the problem of fake news detection and the associated user classification task through machine learning and natural language processing and how our work differs from them and contributes to the progress in this field. In Section 3, we describe how we extend the CoAID dataset and we explain the gold standard for the

Spreader and Checker classification challenge. In Section 4, we explain our approach and our deep learning model. In Section 5, we report the experimental results we achieved in Spreader and Checker classification when applying our architecture on the CoAID extended dataset. In Section 6, we discuss the results obtained with our approach and we explain the choices made for baseline comparison and linguistic model. Finally, we conclude with insights and planned future works in Section 7.

2. Related Work

Since the 2016 U.S. presidential election, the spread of online misinformation on social media platforms such as Twitter and Facebook has produced many publications in this field [10–16]. In [10], Alcott et al. described fake news as articles that are verified as false with certainty and that mislead readers. An example of fake news is an affirmation of a false birth place. This information is verified as false thanks to the data from public registries. In [11], Shu et al. described fake news from a data-mining perspective. In this survey, they explained the existing differences in fake news and related data from traditional media, having stronger psychological and sociological foundations, and from social media platforms, mainly driven by malicious accounts and echo chambers. Once collected, these fake news have been grouped by their textual content and by their social context. In fact, they are characterized by news checked as false, by the use of a specific linguistic style, by their support or denial expressed explicitly beside the news content, and finally by their diffusion behavior through the social community. In [12], Lazer et al. suggested that social media platforms and their content diffusion mechanism are natural habitats for fake news. They advised researching a solution to this problem by creating bot (software-controlled account) and automatic content detection tools to support human supervision to avoid either government or corporate censorship. In [13], Stella et al. addressed the problem of bot detection and effect on social media communities in the case of the Catalan referendum for independence in 2017. They explored the social graph metrics, such as the source and destination of messages between groups as well as sentiment analysis. They found that social bot and humans have different behaviors and that the former tends to stimulate inflammatory reactions in humans. The approaches in the work of Grinberg et al. [14], Guess et al. [15], and Pennycook et al. [16] were all lead by social media user characteristics and related social graph metrics in the context of political elections.

In parallel, the viral content diffusion mechanism through social media has been studied to understand recurrent patterns [7,17–19]. In [17], Shao et al. developed Hoaxy, a monitoring tool on Twitter to understand differences in the diffusion behavior between fake news and fact-checked news. They discovered that social media bots were at the core of the diffusion network and that fact-checked news affects just the peripheral of the same network. Their work is completely based on social media graph metrics such as in and out degree, PageRank, and network diffusion steps. In [18], Dhamal found that exploiting highly influential nodes of a social network community to spread information on a multiple phase scenario does not increase the diffusion effectiveness, while using lesser influential nodes in subsequent phases keeps the pace of the diffusion process high. In [19], Goyal et al. described how social media users are influenced by neighbors in performing actions such as sharing news. They developed a mathematical model based on social media graphs to predict the probability that information is spread through certain nodes of the social media community. In [7], Vosoughi et al. analyzed true and fake news diffusion behavior on Twitter, and they found that social media bots spread true and fake news at the same rate, implying that humans have a major contribution to the phase of cascading the distribution of fake news.

At the same time, other research projects analyzed the impact of network topologies and influence characteristics of certain nodes of the social media communities in the information diffusion mechanism [20–23]. In [20], Zhang et al. developed a new metric called social influence locality to compute the probability that an information is spread by a node in a social media graph based on the behavior of surrounding nodes. They reinforced

the concept that not only the typologies of connection between nodes but also how these nodes behave independently are important. In [21], Guo et al. compared the impact of major influences in a social network graph between global influencers and local influencers, finding that the local ones have an higher probabilistic footprint on a node action. In [22], Mansour et al. highlighted the role played by interpersonal influences between people sharing the same experience in the context of information spreading online. This result suggests that the analysis of users' features contributes to a better precision in prediction of users' sharing action on social media platforms. In [23], Borges et al. investigated what motivates users to share viral communications on the web communities. They observed that users involved in the sharing action of viral contents prefer not to participate in the discussion of the content itself. They classified users in three categories based on their reactions to viral contents. These categories are heavy, social-driven, and search-driven. Heavy users are impacted by content meaningfulness because they interact and produce content heavily on the social media platforms. Social-driven users interact with content mainly by sharing it without adding information, while search-driven does not interact or share the content. The last two categories are more interested in the impact that the information has on how the surrounding users perceive them as a person rather than the meaningfulness of the news content.

Similarly, the linguistic tools and the machine learning methods adopted to extract information from social media posts and text in general have seen an exponential increment in computational power and effectiveness. Since 2018, when the BERT model by Google [24] and the transformer-based architecture combined with the attention mechanism by Vaswani et al. [8] were published, the NLP (Natural Language Processing) methodology has been applied to the misinformation field to inspect this phenomenon [5,25]. In [25], Stieglitz et al. found a positive relationship between the quantity of words containing both positive and negative sentiments rather than the neutral tweets and the probability the social media post will be shared. This result means that sentiment is also spread through social media networks alongside the content of the news itself. In a political context, this concept is validated by the work of Jiang et al. [5].

As for linguistic signals, the user's personality also has an impact on the action of sharing fake news. A number of researchers have employed those features to find the relation between personality traits and the use of social media [26–28]. In [26], Burbach et al. developed an agent-based simulation of a social media interaction. They created these agents modeling answers given by an online questionnaire. The information retrieved were about age; gender; level of education; dimensions of personal social network; and the personality scores from the Five Factor Model, the Dark Triad, and Regulatory Emotional self-efficacy. They used Netlogo, https://ccl.northwestern.edu/netlogo/ (accessed on 16 January 2021), to create the virtual environment and to test the interaction of agents and the diffusion of fake news inside the network. They found that social media graphs, the number of interconnections, and the centrality of nodes have a greater impact than personality scores. Even if this project was tested in a simulated scenario, it suggests that the solution to the problem of fake news diffusion comes from a multi-facets approach both from the psychology of the users and from the structure of the social networks. In [27], Ross et al. described how the user's personality changes their behavior during an interaction with Facebook. Similarly, in [28], Heinstrom et al. described how personality dimensions influence the information diffusion in social media platforms. In this field, Giachanou et al. developed a user-centered CNN model to deal with misinformation spreaders and fact checkers [29]. They developed a multi-input CNN with linguistic features associated with personality traits from the Five Factor Model and the LIWC (Linguistic Inquiry Word Count) [30] dictionary. Their model is word based, and it uses the 300-dimensional pretrained GloVe embeddings [31] to transform textual tweets into a 2D embeddings matrix. These embeddings are the input of a convolutional layer; then, they are processed to compute personality traits and finally merged with manual extracted LIWC features. This approach is innovative because it uses both the personality traits of a

user and their linguistic patterns in the context of fake news. It also proposes a solution that leverages the actions and the motivations of the social media users. On the other hand, this work presents some major limitations; in fact, the computed personality traits have not been validated with a ground truth dataset or with the support of a questionnaire so this initial error is further spread in the successive layers of their neural network. This research project has also some drawbacks in the labeling procedure because it heavily relies on the presence of specific words associated with fact checks or false claims such as hoax, fake, false, fact check, snopes, politifact leadstories, and lead stories, while tweets are labeled as fake if they are a retweet of original fake news. Even if this labeling procedure is manually checked over five hundreds tweets, it is not fully error proof. In addition, stance classification to assess the support or denial of the fake news is not considered at all. Finally, the final number of users analyzed is less than three thousands, and the final f_1 is below 0.6, meaning that the binary classifier has room to be improved.

According to the research projects listed so far, this field of investigation is split into two macro areas. The first one detects fake news contents with the adoption of natural language processing models. The second area monitors the social network topologies to compute how the misinformation spreads among social media users. According to the RQ1, our work inspects the intersection of these two fields with the creation of a linguistic model, focused on COVID-19, to classify fake news spreaders and real news checkers increasing the recall, precision and f_1 of the existing baseline by Giachanou et al. [29]. In fact, Limeng and Dongwon [9] already labeled fake news and real news in their CoAID dataset. We collected users sharing misinformation about COVID-19 from the CoAID dataset, and we downloaded the related Twitter timelines they authored. We transformed this source of information into user embeddings, encoding their tweets, and we exploited their linguistic signals for classification. We released our dataset on Figshare, https://doi.org/10.6084/m9.figshare.14392859 (accessed on 7 May 2021), and the code of our Fake News Spreader Classifier on Github, https://github.com/D2KLab/stopfaker (accessed on 7 May 2021). In addition, we built the RF Fake News Spreader Classifier, a random forest model that exploits a list of features from each Twitter account reaching scores comparable to the ones obtained with the linguistic model. We also developed another deep learning model that receives both tweet embedding and Twitter information as inputs, obtaining lower results in precision, recall, and f_1 with respect to the baseline by Giachanou et al. [29]. In the following section, we describe how we collected the data and how we created the gold standard for this research field to answer our RQ2.

3. Dataset and Gold Standard Creation

The CoAID dataset [9] contains two main resources. The first one is a table storing information about fake news and real news about COVID-19 such as the news URL, the link to the fact checking agency that checked it, the title, the content, the abstract, the publish date, and keywords, as listed in Table 1.

Table 1. Feature descriptions of the news table and user engagement in the CoAID dataset.

Type	Features
News Information	ID, Fact-checking URL, Information URLs, Title, Article title, Content, Abstract, Publish date, Keywords
User Engagement: tweets	ID, Tweet ID

The second one is a list of tweet IDs containing fake and real news and a masked reference ID of the related author. Tweet IDs are divided into four categories: fake and real claims, and fake and real news. The former are just opinions with no URLs inside, and the latter, instead, have an explicit URL redirecting to news. We decided to work with the last ones because we need both the content of the article and the content of the tweet to perform the stance classification. The CoAID dataset includes 4251 fact-checked news and 296,000

related user engagements about COVID-19. The checked fake news are contents that have already been demonstrated as false by fact-checking agencies. In this project, six of these agencies are considered: LeadStories https://leadstories.com/hoax-alert/ (accessed on 13 November 2020), PolitiFact https://www.politifact.com/coronavirus/ (accessed on 13 November 2020), FactCheck.org https://www.factcheck.org/fake-news/ (accessed on 13 November 2020), CheckYourFact https://checkyourfact.com/ (accessed on 13 November 2020), AFP Fact Check https://factcheck.afp.com/ (accessed on 13 November 2020), and Health Feedback https://healthfeedback.org/ (accessed on 13 November 2020). The publishing dates of the collected information range from 1 December 2019 to 1 November 2020. We used tweets containing fake news or real news URLs, and from them, we extracted the social media users who authored them. The CoAID dataset lists the real tweet ID, but for privacy constraints, the user ID of each author is masked, so we need to query the Twitter API to retrieve them. We built an extended version of CoAID by retrieving each user's entire timeline from 1 December 2019. Our extended version of the CoAID dataset is publicly available on Figshare https://doi.org/10.6084/m9.figshare.14392859 (accessed on 13 November 2020), and it is one of the two main contributions of this research project along with the linguistic model to classify fake news spreaders. The retrieval pipeline is described in Figure 1. We collected 11,465 users with an average timeline of 2012 tweets per user. Our text preprocessing phase and data cleaning includes an initial phase of URL extensions because the downloaded tweets contain the Twitter shortened version of the original posted links. As an example, the shortened URL https://t.co/3g8dLgoDOf (accessed on 13 November 2020) has to be extended to https://www.dailymail.co.uk/health/article-9225235/Rare-COVID-arm-effect-leaves-people-got-Modernas-shot-itchy-red-splotch.html (accessed on 13 November 2020). The now extended link was searched inside the CoAID dataset, and if there was a match, we performed the stance detection using the text contained in the original tweet and the abstract of the news from CoAID as input. The stance classification model is an adapted version of the one by Aker et al. [32] in the context of our use case scenario. It is a word-based Random Forest that features Bag of Words, Part of Speech Tagging, Sentiment Analysis, and Named Entity Recognition to classify the source tweet with respect to another one. The entire pipeline was further tuned to work with pretrained multilingual BERT embeddings by the Gate Cloud community, https://cloud.gate.ac.uk/ (accessed on 13 November 2020), and the source code is available in a publicly accessible repository, https://github.com/GateNLP/StanceClassifier (accessed on 13 November 2020). The stance classification output defines whehter the tweet supports, denies, queries, or comments on the linked news. We discard the query and comment cases while counting support and deny ones. If a user supports more fake news than real news, they are labeled as a spreader, and for vice versa, they are labeled as a checker. In the case of an equal number for real and fake news, the user is discarded. The pipeline describing this process is presented in Figure 2. The stance classification algorithm avoids labeling a user as a spreader while they try to refute the fake news spotted. After the data retrieval, data cleaning, and user labeling, we obtained an extended version of the original CoAID dataset to be used as a gold standard. The statistics of this dataset are listed in Table 2. There are 5333 spreaders and 6132 checkers, with an average of 19 tweets supporting fake news per spreader and 55 tweets supporting real news per checker.

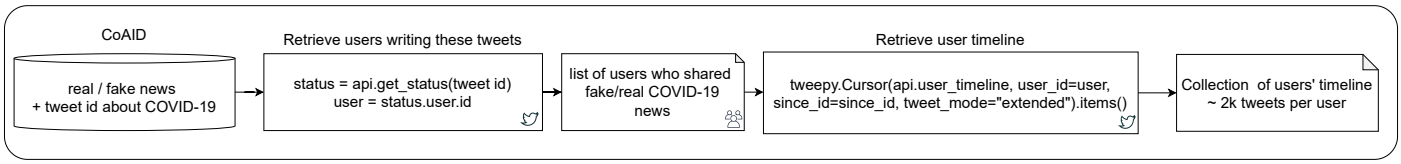

Figure 1. Spreader and checker timeline retrieval extending the CoAID dataset.

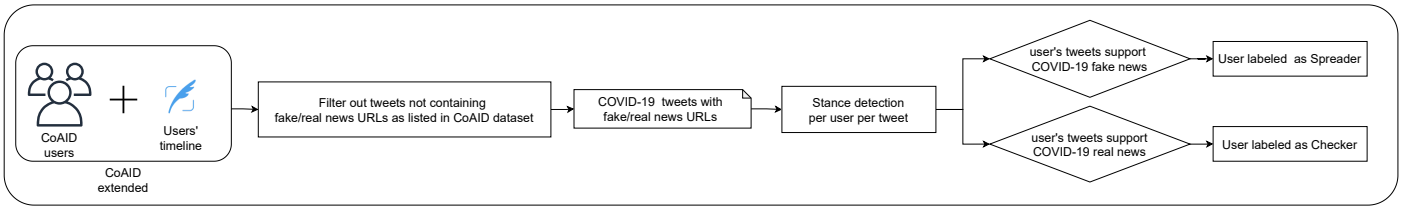

Figure 2. Pipeline to label each user as a spreader or a checker.

Table 2. The CoAID extended dataset statistics.

Total Number of Users	**11,465**
Spreaders	5333
Checkers	6132
Average number of tweets per user	2012
Total number of tweets	23,068,006
Average number of tweets supporting Fake News per Spreader	19
Average number of tweets supporting Real News per Checker	55

We created the gold standard dataset for the classification of users sharing misinformation about COVID-19. The extended version of the CoAID dataset presents a list of mapped user IDs for privacy concerns, the list of real tweet IDs as retrieved from Twitter, and the label classifying the tweet author as a spreader or checker. Five randomly selected rows from the CoAID extended dataset are listed in Table 3. This gold standard answers the RQ2 listed in Section 1.

Table 3. The CoAID extended gold standard dataset used for fake news spreader classification. The user mapped ID is a transformation of the original Twitter user ID to preserve privacy. Tweet ID represents the numerical identifier of the tweet as given by the Twitter platform. The third column represents whether a user is a spreader or a real news checker.

User_MAPPED_id	**Tweet_id**	**Label**
2442	1340854864562311168	1
8885	1346408723330314241	0
6260	1367096980762226688	1
10728	1285659580677193734	0
1956	1352412905199681538	1

4. Approach and Contribution

We created a linguistic model based on a sentence level attention mechanism enhanced by a neural network architecture that differentiates real news checkers and fake news spreaders. The extended version of the CoAID dataset provides the gold standard on which we test our model in three different configurations, and we compared it with the work by Giachanou et al. [29]. We developd a spreader and checker classifier with a text-based linguistic model. In the following, we outline how we process the input batch of tweets and how we develop the stacked neural network. We describe a stacked neural network as a combination of publicly available neural network architectures in which the features are extracted at an intermediate layer of the network and then concatenated together to form a larger feature set. This approach is involved both in the sole text model and in the ensemble with social media metrics used as a comparison model.

4.1. Tweet Embeddings with Transformers

We structure the CoAID extended dataset as a collection of tweets paired with their authors. This data format manifests our attention towards a heavily user centered model. This collection of raw textual tweets, batched per user, represents the written production

of an author. We represent the features of a single tweet transforming its text into a sentence embedding as illustrated in Figure 3. We perform this operation through the multi-headed attention layers [33] in the BERT encoder [24] in Figure 4. Equation (1) describes how the attention mechanism works. Given a sequence of n d-dimensional vectors $x = x_1, \ldots, x_n \in R^d$ and a query vector $q \in R^d$, the attention layer parametrized by $W_k, W_q, W_v, W_o \in R^{dxd}$ computes the weighted sum in Equation (1). W_k represents the matrix of *key weight* vectors, while W_q is the matrix of *query weight*, then W_v is the *value weight* matrix, and finally the W_o is the *output* matrix of weight by which the concatenated attention head are multiplied. The training phase of these four matrices is described in detail in [33]. In self-attention, every x_i is used as the query q to compute a new sequence of representations. Each attention head, A in the equation, is composed of the four W matrices (W_k, W_q, W_v, and W_o) that are learnt during training. W_o and W_v are elements of the weighted average of word vectors, and Wq and Wk are involved in computing the α_i weights. Equation (2) computes the multi-headed attention, M in the equation, where N_h is an independently parameterized attention layer applied in parallel to obtain the final result.

$$\begin{aligned} A_{W_k,W_q,W_v,W_o}(\mathbf{x}, q) &= W_o \sum_{i=1}^{n} \alpha_i W_v x_i \\ \alpha_i &= softmax(\frac{q^T W_q^T W_k x_i}{\sqrt{d}}) \end{aligned} \tag{1}$$

$$M(\mathbf{x}, q) = \sum_{h=1}^{N_h} AW_k, W_q, W_v, W_o(\mathbf{x}, q) \tag{2}$$

$$u_j = \max_{1 \le i \le 768} c_{ij} \tag{3}$$

The resulting batch of embeddings is an intermediate synthesis of a user's textual production. The final combination step is described in Section 4.2. Thanks to the tweet transformation from text to tweet embedding, we obtained a sentence level representation, while the max pooling 1d, described in Equation (3), creates a user embedding u to be processed for the user classification task. In Equation (3), the initial matrix of tweet embeddings is C = $(c_{ij})_{1 \le i \le m, 1 \le j \le 768}$, where m is the total number of tweets collected for the user in analysis and 768 is the dimensions of each tweet embedding. The original text preserves its complete meaning as given by the author because no data cleaning is performed on the original text. As shown in Figure 3, the tweet is split into words and further into smaller tokens. The BERT-tokenizer is able to handle vocabulary words splitting them into smaller sub-strings. At the end of the tokenization phase, each tweet appears as a string type list of tokens. It is important to add a special CLS token at the beginning of the tweet. This is a custom token used for the classification task. The list of tokens is then processed by the successive twelve encoding layers of the architecture to transform each input token into output word embeddings. A representative encoding layer is displayed in Figure 4. All of the word embeddings but the CLS are discarded. We used the CLS embedding as representative of the tweet as a whole because it has been constructed thanks to the attention mechanism and tuned by the surrounding context. In the case of a short tweet, the surrounding context is the entire sentence. A single token has 768 dimensions, following the optimal configuration of the BERT base model [24]. The number 768 comes from the empirical experiment reported in [24], where the authors suggested it as the best number of features comparing the obtained results in different tasks: General Language Understanding Evaluation, Stanford Question Answering Dataset, Named Entity Recognition, and Multi-Genre Natural Language Inference. We adapted the original version of BERT architecture with the multilingual pretrained embeddings to our specific context through the transformers library by Hugging Face: https://huggingface.co/transformers/model_doc/bert.html (accessed on 22 November 2020). The sentence-level attention mechanism is suited to our scenario, where each word is relevant for social

media post understanding. In particular, we have medium and short sentences; thus, the weight of each word is greater in this context. Once each tweet is transformed into an embedding, the user is represented by the list of all their tweet embeddings from their timelines. Each embedding contributes to the dense feature set given as input to the following layers of FNSC to perform the user classification.

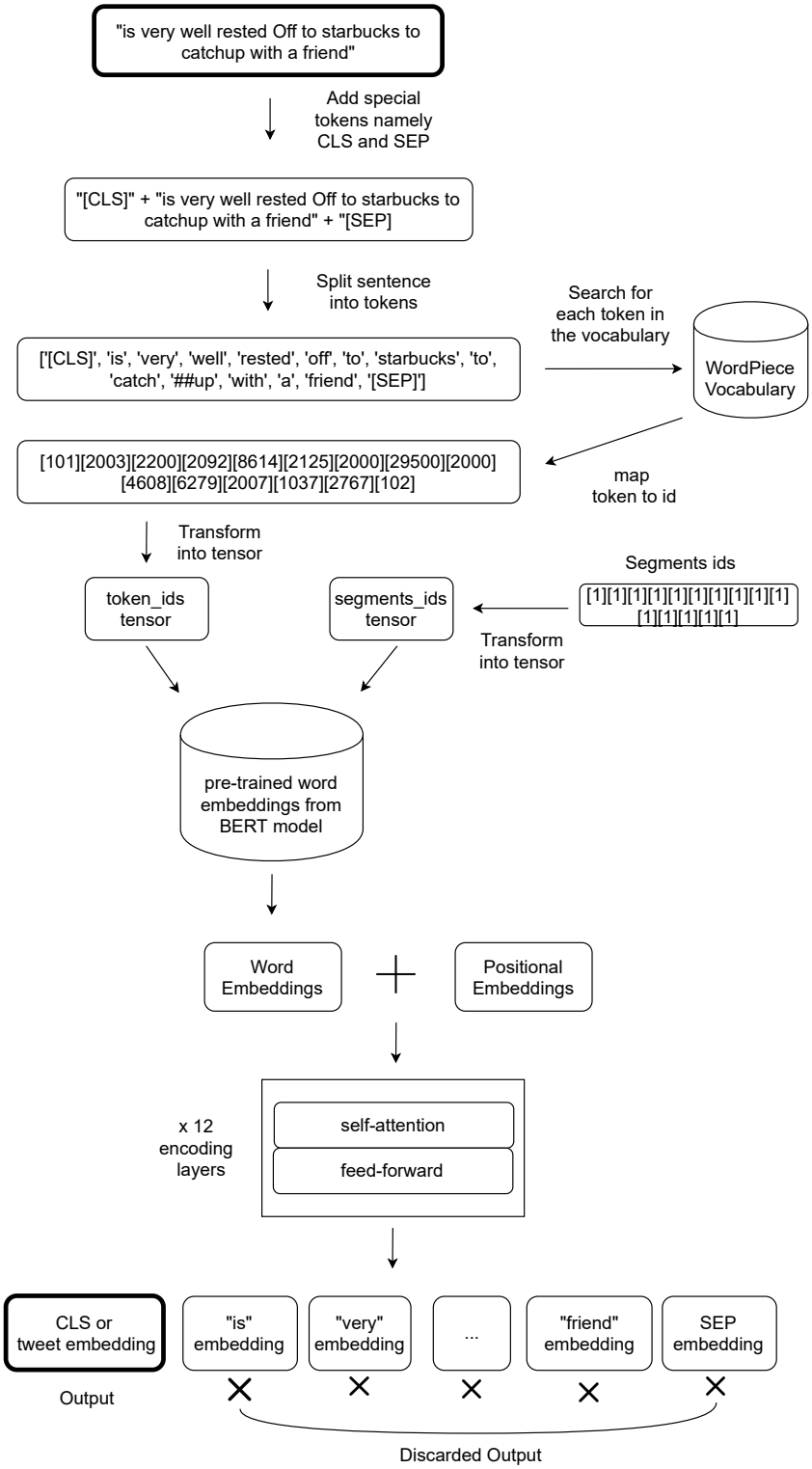

Figure 3. Tokenization and encoding with the transformer. Each tweet in the CoAID extended dataset is processed as shown in the figure. We added a CLS token (classification task special token) at the

beginning of the tweet and the SEP token (separation between sentences) at the end of the tweet. We then split the tweet in tokens. The second part of the splitted words is preceded by ## to tag it as a non standalone word. Tokens are mapped into the ID containing the WordPiece vocabulary and the array so it is transformed into a tensor. We also need a tensor with the same length of token_ids tensor, called segments_ids tensor made of 1s. The segment_ids is useful for dividing tokens belonging to the first sentence (0s) to the second one (1s) when we perform a task that needs two sentences. In our case, we need just a sentence, so we load segments_ids with 1s. We load pretrained embeddings from the BERT model to output word embeddings from our tensors, and we add to them initially random positional embeddings. At the bottom of the figure, there are twelve encoding layers with self attention and a feed forward network inside that encode the input into the final tweet embedding.

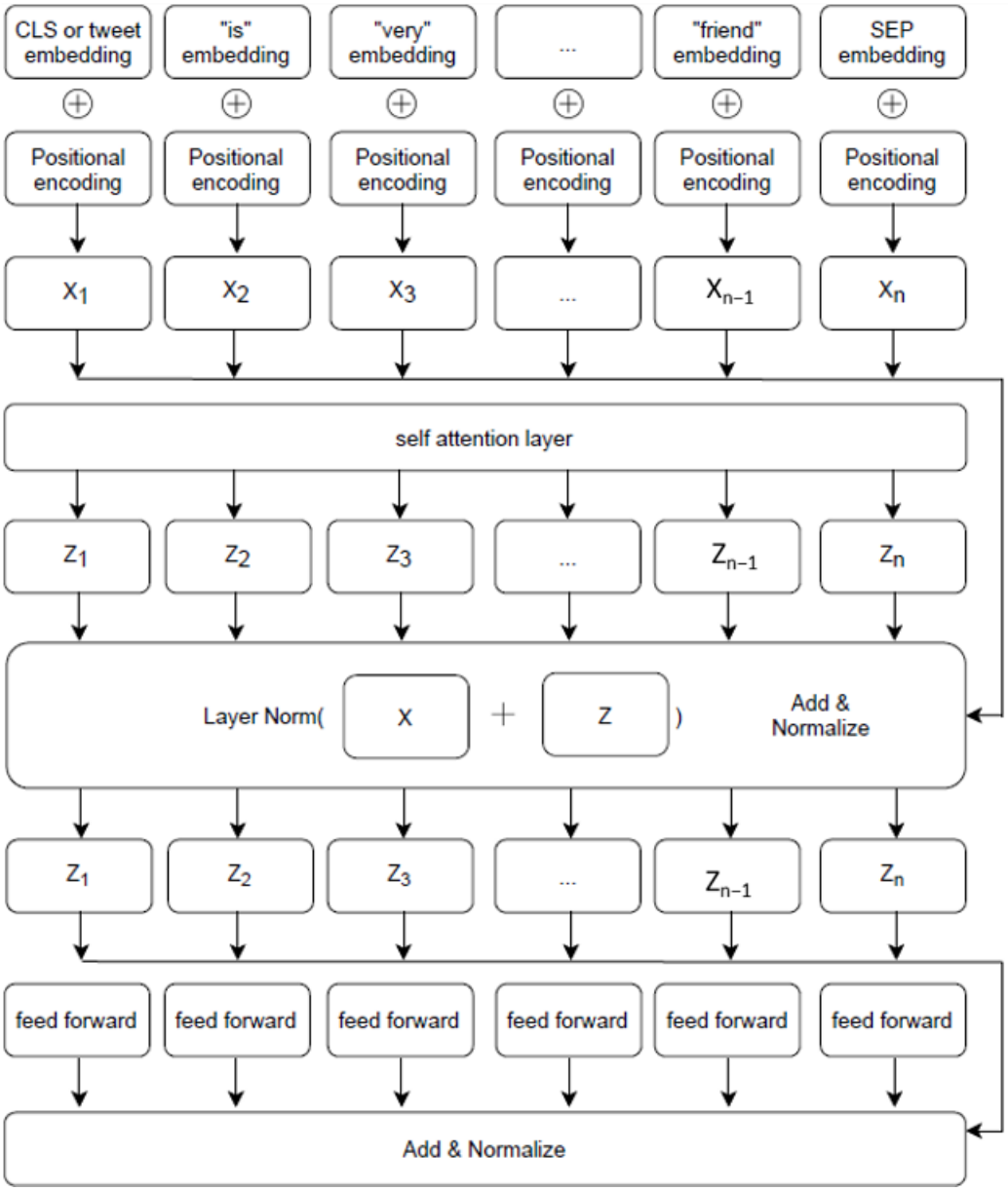

Figure 4. This is a representation of one encoding layer mentioned in Figure 3. There are twelve of these encoding layers in the final architecture. The word embedding of each token passes through these encoding layers, and at the end, we obtain the transformed word embeddings.

4.2. Fake News Spreader Classifier

The Fake News Spreader Classifier, as illustrated in Figure 5, outlines our deep learning model. It receives a batch of tweets, it transforms them into 768-dimensional

arrays, and it stores them in a bidimensional tensor (number of tweets per author, 768). At this point, the intermediate FNSC layer performs a 1d max pooling extraction for each of the 768 dimensions the highest float value. We decided to extract the highest value after empirical tests to make a dense representation of a user's timeline without losing the most specific features of each user. After this stage, the user level embedding is processed by a combination of a Linear Layer paired with Leaky ReLu activation function plus the output layer that is a sigmoid function. The Linear Layer uses the linear function $h_\theta(x) = \sum_j \theta_j x_j = \theta^\top x$ to represent $h(x)$, where $h_\theta(x)$ acts as the linear function family parameterized by θ. The Leaky ReLu function is described in Equation (4) as *f(x)*, and it allows for a small, positive gradient when the unit is not active. This choice improves the performance and speeds up the learning phase. It is also used to avoid the vanishing gradient problem. We have a vanishing gradient in the feed-forward network when we back-propagate the error signal, and it decreases/increases exponentially with respect to the distance from the final layer. Finally, the single neuron with the sigmoid activation function returns a probability that, with a threshold at 0.5, is used to decide the label as 0 or 1 for the final binary classification. We use BCE-loss (Binary Cross Entropy Loss) as a loss function of the architecture. The Equations (5) and (6) presents the Binary Cross-Entropy Loss as used in FNSC to decide how far the prediction is from the expected output, and then, it tunes the neural networks weights with the error back propagation. In particular, y is the label (1 for spreader and 0 for checker) and $p(y)$ is the predicted probability if the sample is a spreader for all N samples in the batch. The formula adds $log(p(y))$ to the loss, the probability of being a spreader. Conversely, it adds $log(1 - p(y))$ to the checker samples. Equation (6) is the contracted form of Equation (5).

$$f(x) = \begin{cases} x & \text{if} \quad x > 0 \\ 0.01x & \text{if} \quad x \geq 0 \end{cases} \tag{4}$$

$$H_p(q) = -\frac{1}{N_{pos} + N_{neg}}[\sum_{i=1}^{N_{pos}} log(p(y_i)) + \sum_{i=1}^{N_{neg}} log(1 - p(y_i))] \tag{5}$$

$$H_p(q) = -\frac{1}{N} y_i \times log(p(y_i)) + (1 - y_i) \times log(1 - p(y_i)) \tag{6}$$

The architecture, presented in Figure 5, is adaptive because it is independent from the number of tweets that a user produces in their timeline. In any case, a user with more than a thousand tweets is better represented than one with few Tweets due to the unbalanced weight of each Tweet in the intermediate user embedding representation.

4.3. Model Optimization

Hyperparameters must be defined both for the encoding phase and the classification phase architectures. In Table 4, regarding the architecture of Figure 3, we set these parameters:

- Pretrained embeddings, the starting point of the original BERT weights to further fine-tune the model based on our data.
- Tokenizer max length, the maximum number of tokens accepted by the BERT-tokenizer.
- Return Tensor, the return tensor format after encoding.
- Hidden Size, the number of neurons in each hidden layer.
- Hidden Layers, the number of layers represented with the self-attention plus feed-forward.
- Attention Heads, this number tunes the self-attention mechanisms described in the work of Vaswani et al. [33].
- Intermediate Size, it represents the number of neurons in the inner neural network of the encoder feed-forward side.
- Hidden Activation Function, it is the nonlinear activation function in the encoder. *GeLu* is the Gaussian Error Linear Unit.

- Dropout Probability, this number represents the probability of training a given node in a layer, where 0 is no training and 1 is always trained.
- Maximum Position Embedding, it is the maximum sequence length accepted by the model.

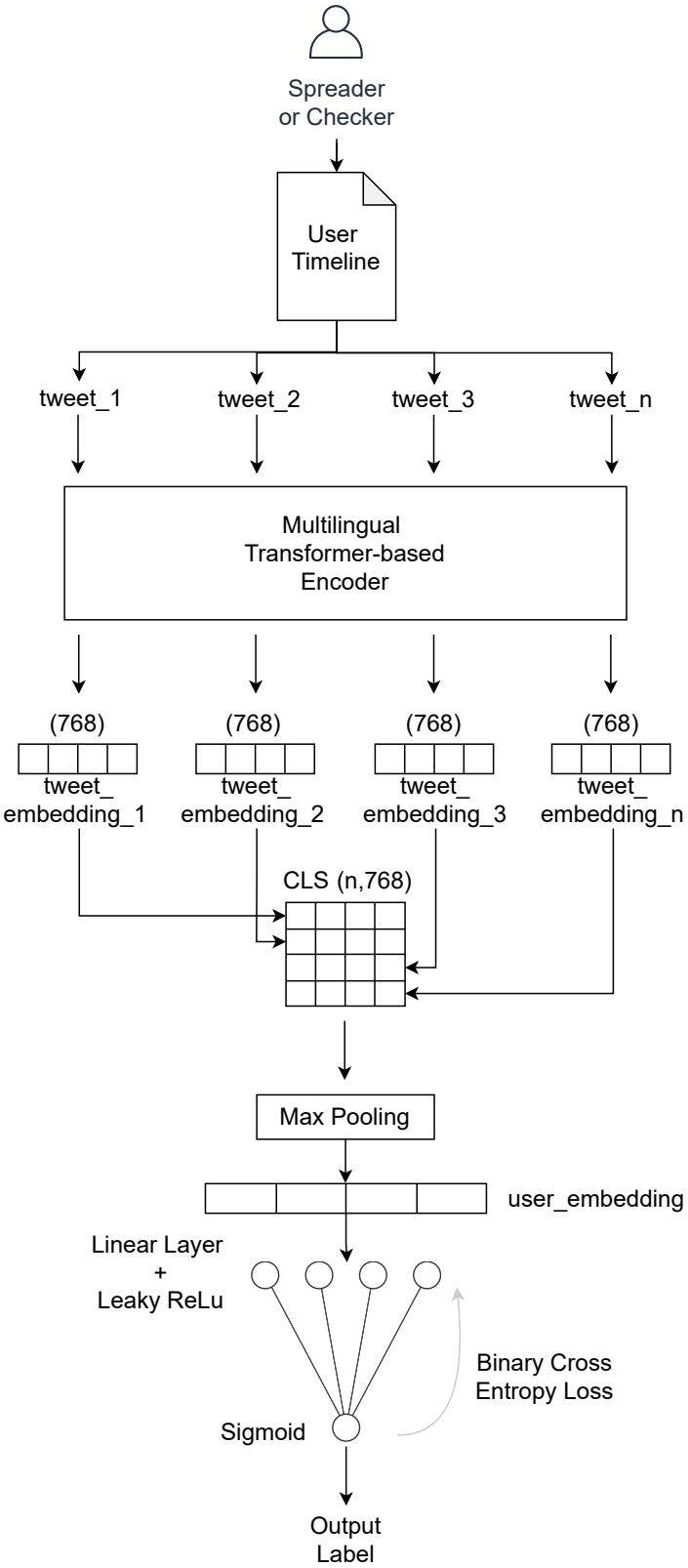

Figure 5. FNSC (Fake News Spreader Classifier) architecture description.

Table 4. Parameters chosen to configure the encoder architecture of Figure 4.

Parameter	Value	Parameter	Value
pre trained embeddings	bert-base-multilingual-cased	tokenizer max length	128
return tensor	pt	hidden size	768
num hidden layers	12	num attention heads	12
intermediate size	3078 (768 × 4)	hidden act	gelu
hidden dropout prob	0.1	max position embedding	512

On the other side, as shown in Table 5, we also have to optimize our neural network architecture with the following parameters:

- Optimizer, it changes the weights of the neurons based on loss to obtain the most accurate result possible.
- Learning Rate, it is the correction factor applied to decrease the loss. Too high values of learning rate lose some details in weights setting, while too low values may lead the model to a very slow convergence.
- Loss Function, it computes the distance between predicted values and actual values.
- Batch Size, it is the number of training examples utilized in one iteration.

Table 5. Parameters to configure the neural network of Figure 5: optimizer, learning rate, and loss function batch size.

Parameter	Value
optimizer	Adam, Adagrad, SGD
learning rate	2×10^{-5}, 1×10^{-2}, 1×10^{-7}
loss function	Binary Cross Entropy Loss
batch size	8, 16, 32

The parameters chosen in our neural network architecture are listed in Table 5. The choices we made are validated empirically.

5. Experimental Results

In this section, we analyze the above model results to assess that we improved the actual state-of-the-art in fake news spreader classification. We compare our model with a previous one by Giachanou et al. [29] as well as different machine learning and deep learning approaches.

The experiment was performed with the CoAID extended dataset (11,465 users and 23 M tweets). In Table 6, we present the results of our model compared with the other configurations and the previous state-of-the-art results, adopting as validation metrics recall, precision, and f_1 score. We used a ten fold cross validation, and we averaged the results obtained for each split. We then performed a ten-fold cross validation to verify the model is effective in each split of the CoAID extended dataset.

Validation metrics, in the binary case, compute how many candidates are well classified with respect to the expected output as described in the following. *tp* (true positive) represents the number of spreaders correctly classified. *tn* (true negative) represents the number of checkers correctly classified. *fp* (false positive) represents the number of checkers classified as spreaders. *fn* (false negative) represents the number of spreaders classified as checkers.

$$p = \frac{tp}{tp + fp} \tag{7}$$

$$r = \frac{tp}{tp + fn} \tag{8}$$

$$f_1 = 2 \times \frac{p \times r}{p + r} \tag{9}$$

As shown by the Equations (7)–(9), where p is the precision metric and r is the recall metric, these metrics give insights about spreader classification effectiveness. This project's goal is primarily to find fake news spreaders because they are the one socially dangerous rather than fact checkers that behave normally. In the binary case, precision and recall are more suitable than accuracy to address this issue. With these premises, we explain the results of each tested model.

Table 6. Precision, recall and f_1 scores computed as a comparison between our proposed Fake News Spreader Classifier model; our RF Fake News Spreader Classifier, that is a Random Forest model exploiting Twitter user's information; the baseline by the work of Giachanou et al.; and a mixed model receiving Twitter user account information and tweet embeddings as input in a stacked neural network. As suggested by the results, we consider the Fake News Spreader Classifier as the most effective in user classification thanks to the overall higher scores.

Model	Precision	Recall	f_1
Fake News Spreader Classifier	0.8042	0.8110	0.8076
RF Fake News Spreader Classifier	0.7977	0.8104	0.804
Giachanou et al. [29]	0.7789	0.7536	0.7660
Mixed Fake News Spreader Classifier	0.7364	0.7430	0.7234

The first line of Table 6 reports the scores obtained with our model as described in Section 4 that uses the sole textual information collected in the user timelines as input.

In parallel, we have in row two the results obtained with the RF Fake News Spreader Classifier, a random forest with 100 estimators, and Gini split criterion with no max depth. Random Forest is an ensemble method that operates by constructing a multitude of decision trees and by outputting the class that is the mode of the classes or mean prediction of the individual trees. In our case, there are two classes: fake news spreader or real news checker. We use 11 as the max_feature parameter, derived from the number of features listed and described in Table 7. We collected the Twitter account information of each user found in the CoAID extended dataset. We transformed the string type features (location and created at) with one hot encoding, while the boolean features (protected, verified, default profile, default profile image) were mapped to 1 or 0. All of the features were then min–max scaled and translated individually such that they are in the given range between zero and one. It is interesting to notice that its metrics are closer to the best performing ones so that it is an index to express the great amount of information contained in social media graph features related to each Twitter user.

The third line in Table 6 shows the final results of the model by Giachanou et al. [29] running on our CoAID extended dataset. We considered their model as the previous state-of-the-art in this field because their work explicitly searches for spreaders and checkers considering user-related features and not just the news. They collect personality traits and psychological signals from the LIWC dictionary of each user.

In the last line, Mixed Fake News Spreader Classifier reports the results we have when we concatenate tweet embeddings and the tabular data containing Twitter user information in the penultimate layer of the FNSC stacked neural network. In the last configuration, the additional information from the users account does not improve the final score. The results presented in Table 6 and the related comments answer the RQ1 listed in Section 1. In fact, we demonstrate that tweet encoding based on transformers and deep learning is effective in fake news spreader classification because they obtain results above 80% in precision, recall, and f_1. It is also better with respect to solutions adopting standard machine learning as the RF Fake News Spreader Classifier described previously.

Table 7. Twitter account user information used for classification with the RF Fake News Spreader Classifier described in Section 4.

Attribute	Data Type	Twitter Attribute Description
location	string	The user-defined location for this account's profile
protected	boolean	When true, indicates that this user has chosen to protect their tweets
verified	boolean	When true, indicates that the user has a verified account
followers count	integer	The number of followers this account currently has
friends count	integer	The number of users this account is following
listed count	integer	The number of public lists that this user is a member of
favourites count	integer	The number of tweets this user has liked in the account's lifetime
statuses count	integer	The number of tweets (including retweets) issued by the user
created at	string	The UTC datetime that the user account was created on Twitter
default profile	boolean	When true, indicates that the user has not altered the theme or background of their user profile
default profile image	boolean	When true, indicates that the user has not uploaded their own profile image and a default image is used instead

6. Discussion

In Section 5 we used precision, recall, and f_1 reported in Table 6 to show that our results improve the current state-of-the-art. Precision registers an increment of 3%, recall registers an increment of 6%, while f1 registers an increment of 4%. We also highlighted the importance of latent information in written text as well as social media graph features. Even if these two sources of information are meaningful for fake news spreader classification tasks, they are not as effective when combined in a single deep learning architecture. This finding suggests that further exploration in this direction should be made. These results and their related considerations answer our RQ1; in fact, we proved that sentence encoding based on transformers and deep learning are the most effective in classifying spreaders of fake news in the context of COVID-19 news. Another important specification has to be made with respect to data retrieval. We were able to collect a greater amount of tweets thanks to the Twitter Academy License, https://developer.twitter.com/en/solutions/academic-research (accessed on 1 February 2021), which was recently released by Twitter for approved research projects. This licence allows us to collect data usually restricted to a standard developer licence account. For privacy concerns, our CoAID extended dataset is released with a mapped user ID so that the anonymity of the authors is preserved. This condition is even more necessary due to the sensible topic and the restricted access we obtained. Another consideration is about the topic of the collected tweets. The original collection of fake and real news as presented in the CoAID dataset are all related to COVID-19 topics, so a broader application with more general topics should be performed to extend the validity of this research project. The extension of the CoAID dataset with the collection of Twitter timelines for more than 12k users, the phase of stance detection to check the support of a user about the fake news they share, and the labelling of the users as fake news spreaders or checkers are the answers to RQ2. The extended CoAID dataset we released is a gold standard for classifying spreaders of fake news in the context of COVID-19 information. Finally, it is important to notice that we do not want to enact a censorship process, instead, acting with the user-centered approach, the consequent step is to activate awareness-raising campaigns towards those users. In parallel with this specification, we affirm that, if the fact checking process requires heavy human supervision, the detection of sensitive users as the target of guidelines suggestions is much easier to scale and automatize.

7. Conclusions and Future Work

We analyzed the socially dangerous issue of misinformation spread from the users perspective. We developed a linguistic model to classify users as fake news spreaders or real news checkers. In this project, we described a language model that processes social media posts written by users; it transforms them into high-dimensional arrays through a transformer-based encoder and max pools them to obtain a user-related high level embedding. This embedding was further used to perform the classification task by the last layer of our FNSC (Fake News Spreader Classifier) stacked neural network. We outperformed the actual state-of-the-art of Giachanou et al. [29] by 4% in f_1 score. We answered RQ1, demonstrating that, even if the social media graph features have a high impact in the fake news spreader classification task, they are less effective than the features extracted from the sole text. In fact, tweet encoding based on transformers and deep learning are effective in classifying spreaders of fake news in the context of COVID-19 news when they are processed in batches as user embeddings. The outcomes of this research project create a new classification parameter in the development of countermeasures against misinformation. The second main contribution of our work is the creation and release of a gold standard for classifying spreaders of fake news in the context of COVID-19 news. The lack of a gold standard in the field of a user-centered classification of fake news spreaders led us to answer RQ2 with the development of the extended version of the CoAID dataset. In future work, we want to address the problem of bot detection in the field of misinformation spread to understand if their semantics are different from real users that spread fake news. We also want to develop a real-time analysis tool to monitor users spreading misinformation on social media by aggregating features from social media metrics, personality metrics, and sentiment in addition to the one related to the text embeddings. We plan to expand the dataset to include other topics in addition to COVID-19 as collected in the original CoAID dataset, and by doing this, we will further expand our CoAID extended dataset. We aim to build automated tools and conversational agents to support human effort in the misinformation contrast and to suggest positive behavior to users who spread fake news in the past or who have characteristics similar to fake news spreaders.

Author Contributions: Conceptualization, S.L. and G.R.; investigation, S.L.; methodology, S.L.; project administration, M.M.; software S.L.; supervision, G.R. and M.M.; validation, G.R.; visualization, S.L.; writing—original draft, S.L.; writing—review and editing, G.R. and M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: CoAID dataset is publicly available at https://github.com/cuilimeng/CoAID (accessed on 16 January 2021). The research output generated by this study is publicly available on FigShare https://doi.org/10.6084/m9.figshare.14392859 (accessed on 7 May 2021). The code developed during this study is publicly available on GitHub https://github.com/D2KLab/stopfaker (accessed on 7 May 2021).

Acknowledgments: Computational resources were provided by HPC@POLITO, a project of Academic Computing within the Department of Control and Computer Engineering at the Politecnico di Torino (http://www.hpc.polito.it (accessed on 21 April 2021)).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

FNSC	Fake News Spreader Classifier
RF	Random Forest
M	Million
k	Thousand
NLP	Natural Language Processing

References

1. Ha, T.; Schensul, S.; Lewis, J.; Brown, S. Early assessment of knowledge, attitudes, anxiety and behavioral adaptations of Connecticut residents to COVID-19. *medRxiv* **2020**. [CrossRef]
2. Cinelli, Q.; Galeazzi, V.; Brugnoli, S.; Zola, Z.; Scala. The COVID-19 social media infodemic. *Sci. Rep.* **2020**, *10*, 16598. [CrossRef] [PubMed]
3. Oshikawa, R.; Qian, J.; Wang, W.Y. A Survey on Natural Language Processing for Fake News Detection. In Proceedings of the 12th Language Resources and Evaluation Conference, Marseille, France, 13–15 May 2020; European Language Resources Association: Marseille, France, 2020; pp. 6086–6093.
4. Islam, M.R.; Liu, S.; Wang, X.; Xu, G. Deep learning for misinformation detection on online social networks: A survey and new perspectives. *Soc. Netw. Anal. Min.* **2020**, *10*, 82. [CrossRef] [PubMed]
5. Jiang, S.; Wilson, C. Linguistic Signals under Misinformation and Fact-Checking: Evidence from User Comments on Social Media. *Proc. ACM Hum. Comput. Interact.* **2018**, *2*, 82. [CrossRef]
6. Glenski, M.; Weninger, T.; Volkova, S. Identifying and Understanding User Reactions to Deceptive and Trusted Social News Sources. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, 15–20 July 2018; Volume 2: Short Papers; Association for Computational Linguistics: Melbourne, Australia, 2018; pp. 176–181. [CrossRef]
7. Vosoughi, S.; Roy, D.; Aral, S. The spread of true and false news online. *Science* **2018**, *359*, 1146–1151. [CrossRef] [PubMed]
8. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is All you Need. In *Advances in Neural Information Processing Systems*; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30.
9. Cui, L.; Lee, D. CoAID: COVID-19 Healthcare Misinformation Dataset. *arXiv* **2020**, arXiv:2006.00885
10. Allcott, H.; Gentzkow, M. Social Media and Fake News in the 2016 Election. *J. Econ. Perspect.* **2017**, *31*, 211–36. [CrossRef]
11. Shu, K.; Sliva, A.; Wang, S.; Tang, J.; Liu, H. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explor. Newsl.* **2017**, *19*, 22–36. [CrossRef]
12. Lazer, D.M.; Baum, M.A.; Benkler, Y.; Berinsky, A.J.; Greenhill, K.M.; Menczer, F.; Metzger, M.J.; Nyhan, B.; Pennycook, G.; Rothschild, D.; et al. The science of fake news. *Science* **2018**, *359*, 1094–1096. [CrossRef] [PubMed]
13. Stella, M.; Ferrara, E.; De Domenico, M. Bots increase exposure to negative and inflammatory content in online social systems. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 12435–12440. [CrossRef] [PubMed]
14. Grinberg, N.; Joseph, K.; Friedland, L.; Swire-Thompson, B.; Lazer, D. Fake news on Twitter during the 2016 US presidential election. *Science* **2019**, *363*, 374–378. [CrossRef] [PubMed]
15. Guess, A.; Nagler, J.; Tucker, J. Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Sci. Adv.* **2019**, *5*, eaau4586. [CrossRef] [PubMed]
16. Pennycook, G.; Rand, D.G. Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 2521–2526. [CrossRef] [PubMed]
17. Shao, C.; Hui, P.M.; Wang, L.; Jiang, X.; Flammini, A.; Menczer, F.; Ciampaglia, G.L. Anatomy of an online misinformation network. *PLoS ONE* **2018**, *13*, e0196087. [CrossRef] [PubMed]
18. Dhamal, S. Effectiveness of diffusing information through a social network in multiple phases. In Proceedings of the 2018 IEEE Global Communications Conference (GLOBECOM), Abu Dhabi, United Arab Emirates, 9–13 December 2018; IEEE: New York City, NY, USA 2018; pp. 1–7.
19. Goyal, A.; Bonchi, F.; Lakshmanan, L.V. Learning influence probabilities in social networks. In Proceedings of the Third ACM International Conference on Web Search and Data Mining, New York, NY, USA, 3–6 February 2010; pp. 241–250.
20. Zhang, J.; Tang, J.; Li, J.; Liu, Y.; Xing, C. Who influenced you? predicting retweet via social influence locality. *ACM Trans. Knowl. Discov. Data (TKDD)* **2015**, *9*, 1–26. [CrossRef]
21. Guo, J.; Zhang, P.; Zhou, C.; Cao, Y.; Guo, L. Personalized influence maximization on social networks. In Proceedings of the 22nd ACM international conference on Information & Knowledge Management, San Francisco, CA, USA, 27 October–1 November 2013; pp. 199–208.
22. Mansour, O.; Olson, N. Interpersonal Influence in Viral Social Media: A Study of Refugee Stories on Virality. In Proceedings of the 8th International Conference on Communities and Technologies, New York, NY, USA, 26–30 June 2017; pp. 183–192.
23. Borges-Tiago, M.T.; Tiago, F.; Cosme, C. Exploring users' motivations to participate in viral communication on social media. *J. Bus. Res.* **2019**, *101*, 574–582. [CrossRef]

24. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805
25. Stieglitz, S.; Dang-Xuan, L. Political communication and influence through microblogging—An empirical analysis of sentiment in Twitter messages and retweet behavior. In Proceedings of the 2012 45th Hawaii International Conference on System Sciences, Maui, HI, USA, 4–7 January 2012; IEEE: New York City, NY, USA 2012; pp. 3500–3509.
26. Burbach, L.; Halbach, P.; Ziefle, M.; Calero Valdez, A. Who Shares Fake News in Online Social Networks? In Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization, Larnaca, Cyprus, 9–12 June 2019; pp. 234–242.
27. Ross, C.; Orr, E.S.; Sisic, M.; Arseneault, J.M.; Simmering, M.G.; Orr, R.R. Personality and motivations associated with Facebook use. *Comput. Hum. Behav.* **2009**, *25*, 578–586. [CrossRef]
28. Heinström, J. Five personality dimensions and their influence on information behaviour. *Inf. Res.* **2003**, *9*, 9-1.
29. Giachanou, A.; Ríssola, E.A.; Ghanem, B.; Crestani, F.; Rosso, P. The role of personality and linguistic patterns in discriminating between fake news spreaders and fact checkers. In Proceedings of the Applications of Natural Language to Information Systems. Saarbrücken, Germany, 24–26 June 2020; pp. 181–192.
30. Pennebaker, J.W.; Chung, C.K.; Ireland, M.; Gonzales, A.; Booth, R.J. The Development and Psychometric Properties of LIWC2007. 2007. Available online: Http://liwc.net/index.php (accessed on 14 September 2015).
31. Pennington, J.; Socher, R.; Manning, C.D. Glove: Global vectors for word representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1532–1543.
32. Aker, A.; Derczynski, L.; Bontcheva, K. Simple Open Stance Classification for Rumour Analysis. In Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017, Varna, Bulgaria, 2–8 September 2017; INCOMA Ltd.: Varna, Bulgaria, 2017; pp. 31–39. [CrossRef]
33. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2017**, arXiv:1706.03762.