



Article

Illusion of Truth: Analysing and Classifying COVID-19 Fake News in Brazilian Portuguese Language

Patricia Takako Endo ^{1,*}, Guto Leoni Santos ², Maria Eduarda de Lima Xavier ¹,
Gleyson Rhuan Nascimento Campos ¹, Luciana Conceição de Lima ³, Ivanovitch Silva ⁴, Antonia Egli ⁵
and Theo Lynn ⁵

- ¹ Programa de Pós-Graduação em Engenharia da Computação, Universidade de Pernambuco, Recife 50720-001, Brazil; mariaeduarda.xavier@upe.br (M.E.d.L.X.); grnc@comp.poli.br (G.R.N.C.)
² Centro de Informática, Universidade Federal de Pernambuco, Recife 50740-560, Brazil; gls4@cin.ufpe.br
³ Programa de Pós-Graduação em Demografia, Universidade Federal do Rio Grande do Norte, Natal 59078-970, Brazil; luciana.lima@ufrn.br
⁴ Programa de Pós-Graduação em Engenharia Elétrica e de Computação, Universidade Federal do Rio Grande do Norte, Natal 59078-970, Brazil; ivanovitch.silva@imd.ufrn.br
⁵ Business School, Dublin City University, Collins Avenue, D09 Y5N0 Dublin, Ireland; antonia.egli@dcu.ie (A.E.); theo.lynn@dcu.ie (T.L.)
* Correspondence: patricia.endo@upe.br



Citation: Endo, T.P.; Santos, G.L.; de Lima Xavier, M.E.; Nascimento Campos, G.R.; de Lima, L.C.; Silva, I.; Egli, A.; Lynn, T. Illusion of Truth: Analysing and Classifying COVID-19 Fake News in Brazilian Portuguese Language. *Big Data Cogn. Comput.* **2022**, *6*, 36. <https://doi.org/10.3390/bdcc6020036>

Academic Editors: Carson K. Leung and Moulay A. Akhloufi

Received: 25 January 2022

Accepted: 28 March 2022

Published: 1 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Public health interventions to counter the COVID-19 pandemic have accelerated and increased digital adoption and use of the Internet for sourcing health information. Unfortunately, there is evidence to suggest that it has also accelerated and increased the spread of false information relating to COVID-19. The consequences of misinformation, disinformation and misinterpretation of health information can interfere with attempts to curb the virus, delay or result in failure to seek or continue legitimate medical treatment and adherence to vaccination, as well as interfere with sound public health policy and attempts to disseminate public health messages. While there is a significant body of literature, datasets and tools to support countermeasures against the spread of false information online in resource-rich languages such as English and Chinese, there are few such resources to support Portuguese, and Brazilian Portuguese specifically. In this study, we explore the use of machine learning and deep learning techniques to identify fake news in online communications in the Brazilian Portuguese language relating to the COVID-19 pandemic. We build a dataset of 11,382 items comprising data from January 2020 to February 2021. Exploratory data analysis suggests that fake news about the COVID-19 vaccine was prevalent in Brazil, much of it related to government communications. To mitigate the adverse impact of fake news, we analyse the impact of machine learning to detect fake news based on stop words in communications. The results suggest that stop words improve the performance of the models when keeping them within the message. Random Forest was the machine learning model with the best results, achieving 97.91% of precision, while Bi-GRU was the best deep learning model with an F1 score of 94.03%.

Keywords: COVID-19; fake news; health misinformation; Brazilian Portuguese language; exploratory data analysis; machine learning; deep learning

1. Introduction

The Internet is a major source of health information [1–3]. The public consumes health content and advice from a wide range of actors including public health agencies, corporations, healthcare professionals and increasingly influencers of all levels [2,4]. In the last decade, with the rise of social media, the volume and sources of health information have multiplied dramatically with an associated rate of propagation. Health information on social media is not subject to the same degree of filtering and quality control by professional gatekeepers common in either public health or commercial sources and is particularly

prone to being out of date, incomplete and inaccurate [5]. Furthermore, there is extensive evidence that individuals and organisations promote health information that is contrary to accepted scientific evidence or public policy, and in extreme cases, is deceptive, unethical and misleading [6,7]. This is also true in the context of online communications relating to the COVID-19 pandemic [8,9]. Within two months of the disclosure of the first COVID-19 case in Wuhan, China, the Director General of the World Health Organisation (WHO) was prompted to declare: “We’re not just fighting an epidemic; we’re fighting an infodemic” [10]. During the first year of the pandemic, the WHO identified over 30 discrete topics that are the subject of misinformation in the COVID-19 discourse [11].

During the first year of the pandemic, Brazil was one of the global epicentres of the COVID-19 pandemic, both in terms of infections and deaths. From 3 January 2020 to 9 February 2021, the WHO reported 9,524,640 confirmed cases of COVID-19 and 231,534 COVID-19-related deaths in Brazil, the third highest rate in the world after the USA and India [12]. As most governments worldwide, the Brazilian government was blindsided by the rapid transmission and impact of COVID-19. For much of 2020, COVID-19 was a pandemic characterised by uncertainty in transmission, pathogenicity and strain-specific control options [13]. Against this backdrop, the Brazilian government had to balance disease mitigation through interventions such as social distancing, travel restrictions and closure of educational institutions and non-essential businesses with the effects they have on Brazilian society and economy. The success of such strategies depends on the effectiveness of government authorities executing both communication and enabling measures, and the response of individuals and communities [14]. Unfortunately, research suggests significant incongruities between advice offered by Brazilian federal government officials and public health agencies [15–17].

Even before the COVID-19 pandemic, Brazil faced challenges in health literacy levels [18] and increasing distrust in vaccines and vaccination [19]. It is established that humans can be both irrational and vulnerable when distinguishing between truth and falsehood [20]. This situation is exacerbated where truths and falsehoods are repeated by traditional sources of trustworthy information, namely the news media and government, and then shared and amplified by peers via social media. In Brazil, the threat of fake COVID-19 news resulted in the Ministry of Health launching an initiative, *Saúde sem Fake News* [21], in an effort to identify and counteract the spread of fake news. There have been no updates since June 2020. In the absence of adequate countermeasures to stem the rise of fake news and against the backdrop of conflicting communications from federal government and public health agencies, Cardoso et al. [18] described Brazil as “... a fertile field for misinformation that hinders adequate measures taken to mitigate COVID-19”.

The complexity of dealing with communication during a health crisis is quite high, as social media, compared with traditional media, is more difficult to monitor, track and analyse [22], and people can easily become misinformed [23]. Given this context, it is important to develop mechanisms to monitor and mitigate the dissemination of online fake news at scale [24]. To this end, machine learning and deep learning models provide a potential solution. However, most of the current literature and datasets are based on resource-rich languages, such as English, to train and test models, and studies with other languages, such as Portuguese, face many challenges to find or to produce benchmark datasets. Our focus is on fake news about COVID-19 in the Brazilian Portuguese language that circulated in Brazil during the period of January 2020 to February 2021. Portuguese is a pluricentric or polycentric language, in that it possesses more than one standard (national) variety, e.g., European Portuguese and Brazilian Portuguese, as well as African varieties. Furthermore, Brazilian Portuguese has been characterised as highly diglossic, i.e., it has a formal traditional form of the language, the so-called H-variant, and a vernacular form, the L-variant, as well as a wide range of dialects [25,26]. The COVID-19 pandemic introduced new terms and new public health concepts to the global linguistic repertoire, which in turn introduced a number of language challenges, not least problems related to the translation and use of multilingual terminology in public health information and medical research

from dominant languages [27]. Consequently, building models based on English language translation which do not take into account the specific features of the Brazilian Portuguese language and the specific language challenges of COVID-19 are likely to be inadequate, thus motivating this work.

This article makes a number of contributions. Firstly, we provide a dataset composed of 11,382 articles in the Portuguese language comprising 10,285 articles labelled “true news” and 1047 articles labelled “fake news” in relation to COVID-19. Secondly, we present an exploratory data analysis on COVID-19 fake news that circulated in Brazil during the first year of the pandemic. Thirdly, we propose and compare machine learning and deep learning models to detect COVID-19 fake news in the Brazilian Portuguese language, and analyse the impact of removing stop words from the messages.

2. Background

Fake news has been defined both in broad and narrow terms and can be characterised by authenticity, intention and whether it is news at all [28]. The broad definition includes non-factual content that misleads the public (e.g., deceptive and false news, disinformation and misinformation), rumour and satire, amongst others [28]. The narrow definition focuses on intentionally false news published by a recognised news outlet [28]. Extant research focuses on differentiating between fake news and true news, and the types of actors that propagate fake news. This paper is focused on the former, i.e., the attributes of the fake news itself. As such, it is concerned with identifying fake news based on characteristics such as writing style and quality [29], word counts [30], sentiment [31] and topic-agnostic features (e.g., a large number of ads or a frequency of morphological patterns in text) [32].

As discussed in the Introduction, the Internet, and in particular social media, is transforming public health promotion, surveillance, public response to health crises, as well as tracking disease outbreaks, monitoring the spread of misinformation and identifying intervention opportunities [33,34]. The public benefits from improved and convenient access to easily available and tailored information in addition to the opportunity to potentially influence health policy [33,35]. It has had a liberating effect on individuals, enabling users to search for both health and vaccine-related content and exchange information, opinions and support [36,37]. Notwithstanding this, research suggests that there are significant concerns about information inaccuracy and potential risks associated with the use of inaccurate health information, amongst others [38–40]. The consequences of misinformation, disinformation and misinterpretation of health information can interfere with attempts to mitigate disease outbreak, delay or result in failure to seek or continue legitimate medical treatment as well as interfere with sound public health policy and attempts to disseminate public health messages by undermining trust in health institutions [23,41].

Historically, the news media has played a significant role in Brazilian society [42]. However, traditional media has been in steady decline in the last decade against the backdrop of media distrust (due to perceived media bias and corruption) and the rise of the Internet and social media [43]. According to the Reuters Institute Digital News Report 2020 [44], the Internet (including social media) is the main source of news in Brazil. It is noteworthy that Brazil is one of a handful of countries where across all media sources the public prefers partial news, a factor that can create a false sense of uniformity and validity and foster the propagation of misinformation [44]. While Facebook is a source of misinformation concern in most countries worldwide, Brazil is relatively unique in that WhatsApp is a significant channel of news and misinformation [44]. This preference of partial news sources and social media in Brazil has led to significant issues in the context of COVID-19.

From the beginning of the COVID-19 pandemic, the WHO has reported on a wide variety of misinformation related to COVID-19 [11]. These include unsubstantiated claims and conspiracy theories related to hydroxychloroquine, reduced risk of infection, 5G mobile networks and sunny and hot weather, amongst others [11]. What differs in the Brazilian context is that the Brazilian public has been exposed to statements from the political elite,

including the Brazilian President, that have contradicted the Brazilian Ministry of Health, pharmaceutical companies and health experts. Indeed, the political elite in Brazil have actively promoted many of the misleading claims identified by the WHO. This has included statements promoting erroneous information on the effects of COVID-19, “cures” and treatments unsupported by scientific evidence and an end to social distancing, amongst others [45]. These statements by government officials become news and lend legitimacy to them. As vaccines and vaccination programmes to mitigate COVID-19 become available, such statements sow mistrust in health systems but provide additional legitimacy to anti-vaccination movements that focus on similar messaging strategies, e.g., questioning the safety and effectiveness of vaccines, sharing conspiracy theories, publishing general misinformation and rumours, promoting that Big Pharma and scientific experts are not to be trusted, stating that civil liberties and human’s freedom of choice are endangered, questioning whether vaccinated individuals spread diseases and promoting alternative medicine [46–48].

While vaccines and vaccinations are a central building block of efforts to control and reduce the impact of COVID-19, vaccination denial and misinformation propagated by the anti-vaccination movement represents a tension between freedom of speech and public health. Social network platforms have been reluctant to intervene on this topic and on misinformation in general [49], however, there have been indicators that this attitude is changing, particularly in the context of COVID-19 [50]. However, even where there is a desire to curb misinformation by platforms, the identification of fake news and misinformation, in general, is labour intensive and particularly difficult to moderate on closed networks such as WhatsApp. To scale such monitoring requires automation. While over 282 million people speak Portuguese worldwide, commercial tools and research has overwhelmingly focused on the most popular languages, namely English and Chinese. This may be due to the concentration of Portuguese speakers in a relatively small number of countries. Over 73% of native Portuguese speakers are located in Brazil and a further 24% in just three other countries—Angola, Mozambique and Portugal [51]. As discussed earlier, it is important to note that Portuguese as a language is pluricentric and Brazilian Portuguese is highly diglossic, thus requiring native language datasets for accurate classification.

3. Related Works

Research on automated fake news detection typically falls in to two main categories, approaches based on knowledge, and those based on style [20]. Style-based fake news detection, the focus of this article, attempts to analyse the writing style of the target article to identify whether there is an attempt to mislead the reader. These approaches typically rely on binary classification techniques to classify news as fake or not based on general textual features (lexicon, syntax, discourse, and semantic), latent textual features (word, sentence and document) and associated images [20]. These are typically based on data mining and information retrieval, natural language processing (NLP) and machine learning techniques, amongst others [20,52]. This study compares machine learning and deep learning techniques for fake news detection.

There is well-established literature on the use of traditional machine learning for both knowledge-based and style-based detection. For example, naive Bayes [53,54], support vector machine (SVM) [54–58], Random Forest [59,60], and XGBoost [59,61] are widely cited in the literature. Similarly, a wide variety of deep learning techniques have been used including convolutional neural networks (CNNs) [62–65] long short term memory (LSTM) [66,67], recurrent neural networks (RNN) and general recurrent units (GRU) models [66–68], other deep learning neural networks architectures [69–71] and ensemble approaches [63,72,73].

While automated fake news detection has been explored in health and disease contexts, the volume of research has expanded rapidly since the commencement of the COVID-19 pandemic. While a comprehensive review of the literature is beyond the scope of this article, four significant trends are worthy of mention. Firstly, although some studies use a variety of news sources (e.g., [74]) and multi-source datasets such as CoAID [75],

the majority of studies focus on data sets comprising social media data and specifically Twitter data, e.g., [76,77]. This is not wholly unsurprising as access to the Twitter API is easily accessible and the public data sets on the COVID-19 discourse have been made available, e.g., [78–80]. Secondly, though a wide range of machine learning and deep learning techniques feature in studies including CNNs, LSTMs and others, there is a notable increase in the use of bidirectional encoder representations from transformers (BERT) [74,76,77]. This can be explained by the relative recency and availability of BERT as a technique and early performance indicators. Thirdly, and related to the previous points, few datasets or research identified use a Brazilian Portuguese language corpus and a Brazilian empirical context. For example, the COVID-19 Twitter Chatter dataset features English, French, Spanish and German language data [79]. CoAID does not identify its language, but all sources and search queries identified are English language only. The Real Worry Dataset is English language only [80]. The dataset described in [78] does feature a significant portion of Portuguese tweets, however, none of the keywords used are in the Portuguese language and the data is Twitter only. Similarly, the MM-COVID dataset features 3981 fake news items and 7192 trustworthy items in six languages including Portuguese [81]. While Brazilian Portuguese is included, it would appear both European and Brazilian Portuguese are labelled as one homogeneous language, and the total number of fake Portuguese language items is relatively small (371).

Notwithstanding the foregoing, there has been a small number of studies that explore fake news in the Brazilian context. Galhardi et al. [82] used data collected from the *Eu Fiscalizo*, a crowdsourcing tool where users can send content that they believe is inappropriate or fake. Analysis suggests that fake news about COVID-19 is primarily related to homemade methods of COVID-19 prevention or cure (85%), largely disseminated via WhatsApp [82]. While this study is consistent with other reports, e.g., [44], it comprises a small sample (154 items) and classification is based on self-reports. In line with [83,84], Garcia Filho et al. [85] examined temporal trends in COVID-19. Using Google Health Trends, they identified a sudden increase in interest in issues related to COVID-19 from March 2020 after the adoption of the first measures of social distance. Of specific interest to this paper is the suggestion by Garcia Filho et al. that unclear messaging between the President, State Governors and the Minister of Health may have resulted in a reduction in search volumes. Ceron et al. [86] proposed a new Markov-inspired method for clustering COVID-19 topics based on evolution across a time series. Using a dataset 5115 tweets published by two Brazilian fact-checking organisations, *Aos Fatos* and *Agência Lupa*, their data also suggested the data clearly revealed a complex intertwining between politics and the health crisis during the period under study.

Fake news detection is a relatively new phenomenon. Monteiro et al. [87] presented the first reference corpus in Portuguese focused on fake news, Fake.Br corpus, in 2018. The Fake.Br. corpus comprises 7200 true and fake news items and was used to evaluate an SVM approach to automatically classify fake news messages. The SVM model achieved 89% accuracy using five-fold cross validation. Subsequently, the Fake.Br corpus was used to evaluate other techniques to detect fake news. For example, Silva et al. [88] compare the performance of six techniques to detect fake news, i.e., logistic regression, SVM, decision tree, Random Forest, bootstrap aggregating (bagging) and adaptive boosting (AdaBoost). The best F1 score, 97.1%, was achieved by logistic regression when stop words were not removed and the traditional bag-of-words (BoW) was applied to represent the text. Souza et al. [89] proposed a linguistic-based method based on grammatical classification, sentiment analysis and emotions analysis, and evaluated five classifiers, i.e., naive Bayes, AdaBoost, SVM, gradient boost (GB) and K-nearest neighbours (KNN) using the Fake.Br corpus. GB presented the best accuracy, 92.53%, when using emotion lexicons as complementary information for classification. Faustini et al. [90] also used the Fake.Br corpus and two other datasets, one comprising fake news disseminated via WhatsApp, as well as a dataset comprising tweets, to compare four different techniques in one-class classification (OCC)—SVM, document-class distance (DCD), EcoOCC (an algorithm based

on k-means) and naive Bayes classifier for OCC. All algorithms performed similarly with the exception of the one-class SVM, which showed greater F-score variance.

More recently, the Digital Lighthouse project at the Universidade Federal do Ceara in Brazil has published a number of studies and datasets relating to misinformation on WhatsApp in Brazil. These include FakeWhatsApp.BR [91] and COVID19.BR [92,93]. The FakeWhatsApp.BR dataset contains 282,601 WhatsApp messages from users and groups from all Brazilian states collected from 59 groups from July 2018 to November of 2018 [91]. The FakeWhatsApp.BR corpus contains 2193 messages labelled misinformation and 3091 messages labelled non-misinformation [91]. The COVID-19.BR contains messages from 236 open WhatsApp groups with at least 100 members collected from April 2020 to June 2020. The corpus contains 2043 messages, 865 labelled as misinformation and 1178 labelled as non-misinformation. Both datasets contain similar data, i.e., message text, time and date, phone number, Brazilian state, word count, character count and whether the message contained media [91,93]. Cabral et al. [91] combined classic natural language processing approaches for feature extraction with nine different machine learning classification algorithms to detect fake news on WhatsApp, i.e., logistic regression, Bernoulli, complement naive Bayes, SVM with a linear kernel (LSVM), SVM trained with stochastic gradient descent (SGD), SVM trained with an RBF kernel, K-nearest neighbours, Random Forest (RF), gradient boosting and a multilayer perceptron neural network (MLP). The best performing results were generated by MLP, LSVM and SGD, with a best F1 score of 0.73, however, when short messages were removed, the best performing F1 score rose to 0.87. Using the COVID19.BR dataset, Martins et al. [92] compared machine learning classifiers to detect COVID-19 misinformation on WhatsApp. Similar to their earlier work [91], they tested LSVM and MLP models to detect misinformation in WhatsApp messages, in this case related to COVID-19. Here, they achieved a highest F1 score of 0.778; an analysis of errors indicated errors occurred primarily due to short message length. In Martins et al. [93], they extend their work to detect COVID-19 misinformation in Brazilian Portuguese WhatsApp messages using bidirectional long-short term memory (BiLSTM) neural networks, pooling operations and an attention mechanism. This solution, called MIDeepBR, outperformed their previous proposal as reported in [92] with an F1 score of 0.834.

In contrast with previous research, we build and present a new dataset comprising fake news in the Brazilian Portuguese language relating exclusively to COVID-19 in Brazil. In contrast with Martins et al. [92] and Cabral et al. [91], we do not use a WhatsApp dataset, which may due to its nature be dominated by L-variant Brazilian Portuguese. Furthermore, the dataset used in this study is over a longer period (12 months) compared with Martins et al. [92,93] and Cabral et al. [91]. Furthermore, unlike Li et al. [81], we specifically focus on the Brazilian Portuguese language as distinct from European or African variants. To this end, the scale of items in our dataset is significantly larger than, for example, the MM-COVID dataset. In addition to an exploratory data analysis of the content, we evaluate and compare machine learning and deep learning approaches for detecting fake news. In contrast with Martins et al. [93], we include gated recurrent units (GRUs) and evaluate both unidirectional and bidirectional GRUs and LSTMs, as well as machine learning classifiers.

4. Exploratory Data Analysis

The exploratory data analysis (EDA) is focused on COVID-19-related fake content that was debunked between January 2020 and February 2021 and uses the dataset described in Section 4.1. The EDA was performed to inform two research questions:

- Question 1: What are the main themes in fake news related to COVID-19 in Brazil during the first wave of the pandemic and how did they change during the period under study?
- Question 2: Are there textual attributes that characterise COVID-19 fake news?

4.1. Dataset

The relative dearth of Brazilian Portuguese fake news datasets from heterogeneous sources adversely impacts the study of fake news in relation to COVID-19 in Brazil. As such, the first contribution of this study is the development of a new reference dataset for Brazilian Portuguese language fake news items related to COVID-19.

There are a number of Brazilian fact-checking initiatives focused on identifying fake news about COVID-19. These include *Saúde sem Fake News* (<https://antigo.saude.gov.br/fakenews/>, last accessed on 24 January 2022), promoted by the Brazilian Ministry of Health, and *Fake ou Fato* (<https://g1.globo.com/fato-ou-fake/coronavirus/>, last accessed on 24 January 2022), maintained by *Grupo Globo*, one of the most prominent Brazilian media publishers. Unfortunately, these services present challenges when using them as reference datasets for fake news detection. Figure 1a,b presents examples of content published by *Saúde sem Fake News* on their portal and *Fake ou Fato*, respectively. In both instances, the original fake news content is embedded in a new image and labelled. This makes both difficult to read the entire text of, thus compromising the utility of the content for our purposes, also the original raw text is not made available.



Figure 1. Examples of how fake news is presented by (a) the *Saúde sem Fake News* [21] and (b) *Fake ou Fato* [94].

A second source of fake news are expert-based fact-checking organisations. These organisations rely on subject matter domain experts to verify a specific item of news content [20]. *Agência Lupa* (<https://piaui.folha.uol.com.br/lupa/>, last accessed on 24 January 2022), *Aos Fatos* (<https://www.aosfatos.org/>, last accessed on 24 January 2022) and *Projeto Comprova* (<https://projeto comprova.com.br/>, last accessed on 24 January 2022) are amongst the leading expert-based fact-checking organisations in Brazil [86]. Unfortunately, while these organisations check the veracity of the news items and explain why it is considered fake or not, such as *Saúde sem Fake News* and *Fake ou Fato*, they do not publish the original raw text. For the purposes of our dataset construction, data was collected from *boatos.org* (<https://www.boatos.org/>, last accessed on 24 January 2022), a Brazilian independent expert-based fact-checking website. *Boatos.org* was established to collect fake news disseminated online and is updated daily by a team of four Brazilian journalists. Importantly for this study, *boatos.org* established a specific section of their website that focused specifically on COVID-19 fake news and that included the original raw text of the fake news item. To collect the data, we developed a web scraping script (available at GitHub (<https://github.com/rhuancampos/covid-19-rumor>, last accessed on 24 January 2022)) to automatically collect the text-based content of the fake news item debunked by *boatos.org*. The dataset is composed of the following rows: link, date, title, text and classification (1 for fake news). For binary classification, a dataset of both fake news items and non-fake (true) news items is required. True news items were collected from articles published by

G1 (<https://g1.globo.com/bemestar/coronavirus/>, last accessed on 24 January 2022), a reputable and leading Brazilian media portal. Again, we used a web scraping script to collect the true news samples. The dataset is composed of the following rows: link, date, title, text, source and classification (0 for true news).

Due to noise (e.g., text in other languages, text duplication and so on), the data was manually checked and cleaned. The final dataset comprised 1047 fake news items and 10,285 true news items published from 26 January 2020 to 28 February 2021. The entire dataset is available for download at GitHub (<https://github.com/rhuancampos/covid-19-rumor>, last accessed on 24 January 2022).

Figure 2a,b presents the monthly distribution of fake news items and true news items in the dataset, respectively. The number of news about COVID-19 published by G1. In both cases, the dataset peaks in the March–May period.

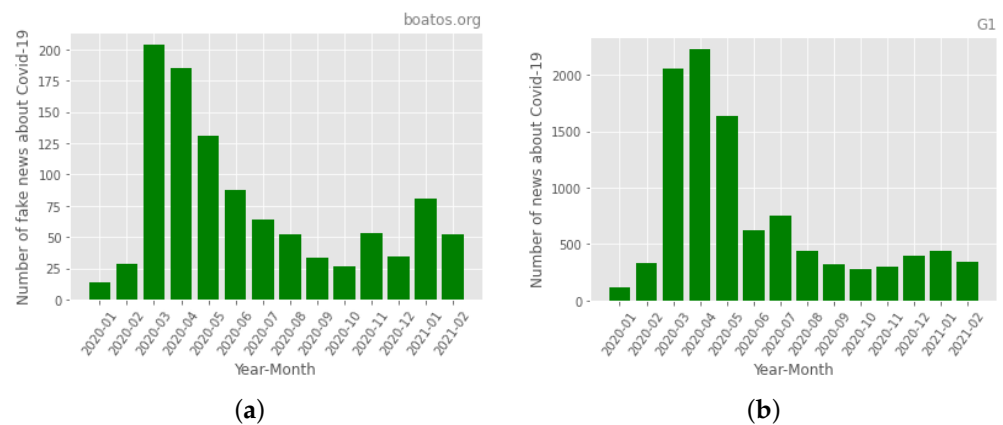


Figure 2. Monthly distribution of (a) fake news items and (b) true news items about COVID-19 in the dataset from 26 January 2020 to 28 February 2021.

Figure 3a,b present the word cloud and the word count for the fake news dataset, respectively. After discounting *virus*, *coronavirus* and *COVID19*, the most common words related to *vacina* (vaccine) and *china*.

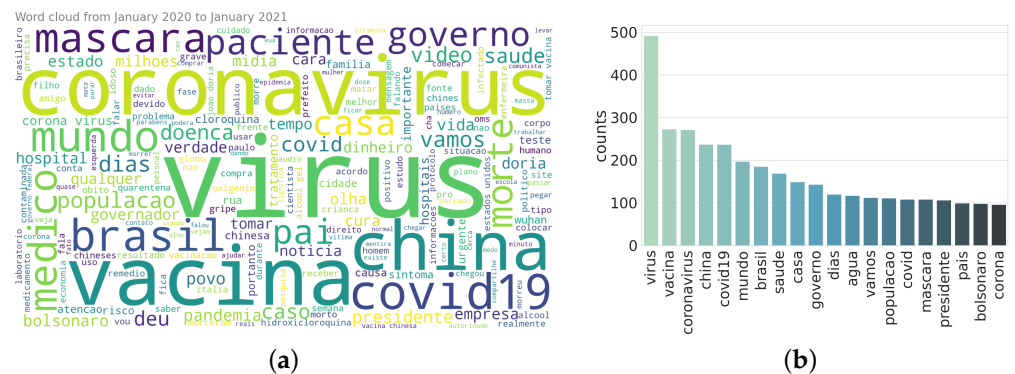


Figure 3. Word cloud (a) and word count (b) for the most popular words in the fake news dataset.

4.2. Fake News Content per Month

Figure 4 shows a bubble graph determined by the most common words per month in the fake news dataset. Unsurprisingly, the words *virus*, *coronavirus*, *COVID-19* and *China* dominate the initial months before *mascara* (mask) and *mortes* (deaths) begin to emerge in April. *Vacina* (vaccine) started to appear more frequently in the fake news from July onwards.



Figure 4. Bubble graph tracking the top 3 words per month featured in the fake news dataset for the period.

The first explicit preventive and mitigation measures for COVID-19 emerged between January and March 2020, particularly in response to the confirmation of the first case in Brazil in late February and the first death in the first half of March. In this period, fake news focused on the origin of the coronavirus including xenophobic content against China and about COVID-19 prevention, e.g. the usage of alcohol-based hand sanitizers. Of the examples below, the first message states that Chinese exotic food was the origin of coronavirus; the second is a recipe for homemade alcohol-based hand rub.

Origin of COVID-19. *Isso é a China! Atenção Cenas Fortes! A mesma que alimentação exótica, deu início a H1N1 a SARS dona do coronavírus. Cenas Fortes para amantes de PETS, lá eles são comida. Mas não julguem existe escassez de agricultura, pecuária e outros. Sopa de morcego e cobra rançosa. Seria a fonte da nova Epidemia Coronavírus. Vitamina C, hidratação, garganta úmida, mãos limpas. Ajuda!*

Homemade alcohol-based hand rub recipe. **Álcool Gel* Ingredientes: 2 folhas de gelatina incolor sem sabor; 1 copo de água; 12 copos de álcool 96°. *Modo de fazer:* Dissolva a gelatina em água quente; Espere esfriar um pouquinho; Acrescente o álcool 96° e misture bem. Está pronto o álcool de 72° a 75°. *ATENÇÃO* Quem for fazer o preparo, lembre-se de manter o álcool líquido afastado do fogo, no momento em que estiver dissolvendo a gelatina.*

In April 2020, fake news suggesting masks originating from China or India were contaminated with COVID-19 began to circulate in Brazil. This was debunked following the first WHO recommendation on the use of masks (<https://apps.who.int/iris/handle/10665/331693>, last accessed on 24 January 2022).

Masks are contaminated with COVID-19. *Segundo alerta da Organização Mundial da Saúde, máscaras vindas da Índia e da China estão apresentando um alto grau de contaminação por coronavírus. Essas máscaras são produzidas às pressas, em lugares impróprios e sem o mínimo controle de cuidado higiênico. Neste exato momento, pessoas que pensam estar seguras por usar a máscara de proteção estão na verdade correndo um grande risco de contaminação. A recomendação crucial da OMS é que todos usem máscaras esterilizadas para que o vírus não se espalhe ainda mais, agravando de maneira descontrolada o avanço da pandemia.*

At the end of April the Amazonas health system collapsed, and as a consequence, the number of daily burials increased from 30 to over 100 (<https://www1.folha.uol.com.br/cotidiano/2020/04/enterros-triplicam-e-cemiterio-de-manaus-abre-valas-comuns-para-vitimas-do-coronavirus.shtml>, last accessed on 24 January 2022). This period was referred to as

“funerary collapse” (<https://noticias.uol.com.br/colunas/leonardo-sakamoto/2020/04/21/negacionistas-do-coronavirus-acham-que-colapso-funerario-em-manaus-e-fake.htm>, last accessed on 24 January 2022). As a result, fake news emerged suggesting that the number of deaths resulting from COVID-19 were fake, that people were being buried alive and coffins were being filled with stones and wood instead of people.

Coffins with stone and wood. *“Oi, Hernandes. Deixa eu falar uma coisa pra você. A última notícia, né, que a Globo não vai falar, mas aqui em Minas tá acontecendo um caso muito engraçado, principalmente, lá em BH. Você sabe quem que é o prefeito de BH, né? Lá tá infestado de coronavírus. Aí, estão enterrando um monte de gente com coronavírus lá em BH. Assim, a própria família tá enterrando, porque não tá dando tempo dos coveiros enterrar, pra não ter aglomeração de pessoas, sabe? Aí, que aconteceu: mandaram ir lá e arrancar todos os caixão para poder fazer o exame, para ver se é coronavírus mesmo. Sabe o que tem dentro do caixão? Pedra e madeira. Um monte de caixão cheio de pedra e madeira. Palhaçada, não? Ah, fala a verdade esses político. Você sabe quem é o prefeito de lá, né? De BH? Ah, gente, por que a Globo não filma isso? Mas já que vai espalhar aí, o povo que tá filmando”*

In June 2020, Agência Nacional de Vigilância Sanitária (Anvisa), the National Health Surveillance Agency authorised the start of COVID-19 vaccine tests in Brazil (<https://agenciabrasil.ebc.com.br/saude/noticia/2020-06/testes-com-vacina-de-oxford-contra-COVID-19-comecam-em-sao-paulo>, last accessed on 24 January 2022). Fake news surrounding the term *vacina* increased during this time (<https://butantan.gov.br/noticias/governo-de-sao-paulo-e-butantan-dao-inicio-aos-testes-da-vacina-contra-coronavirus>, last accessed on 24 January 2022). In this period, fake news content was focused on the Chinese vaccine, including xenophobic comments.

Xenophobic message about Chinese vaccine. *#JoãoPinóquioDória promove a vacina #chingling. A Primeira Ministra da #Austrália também promove a Vacina chinesa. Mas ela nem a toma, só finge tomar ... será por quê ?? *E vc, vai tomar essa vacina do cão ?!?*

From August to December 2020, fake news content consistently revolved around the vaccine, particularly concerning its side effects and linking the vaccine to death (see Figure 4).

Vaccines kill people. *Uma família inteira tomou a vacina chinesa, os 3 filhos morreram a mulher está na UTI. Que tristeza! Confirmado pela própria Flavia Molloy, o caso da família q tomou a vacina chinesa em SP, e matou os 3 filhos e a mãe ta na UTI.*

In January 2021, a large number of fake news items regarding the commencement and efficacy of vaccination in Brazil appeared. These included items suggesting that the vaccine did not provide immunity against COVID-19 and that the adverse side effects were significant.

Side effects of the vaccine. *HOSPITAL LOTADO DE PACIENTES QUE TOMARAM A VACINA DO COVID 19 COM EFEITOS ADVERSOS GRAVES. O EXPOSITOR DO SISTEMA. Ou seja, as venenocinas estão infectando MAIS AINDA as pessoas, do que o próprio vírus.*

Despite previous studies suggesting that the political elite in Brazil played a significant role in the dissemination of fake news relating to COVID-19, President Bolsonaro did not appear in the fake news dataset. In contrast, *governador* (governor), referring to the Governor of the State of São Paulo, João Dória, appeared prominently. João Dória was considered one of President Bolsonaro’s main political opponents. Doria was active in promoting vaccination, and vaccination commenced in in São Paulo in advance of the national campaign (<https://www.saopaulo.sp.gov.br/noticias-coronavirus/government-o>

[f-sp-starts-vaccination-against-coronavirus-on-january-25/](#), last accessed on 24 January 2022). Under Dória, São Paulo financed the production of Coronavac, a vaccine developed by the Chinese biopharmaceutical company, Sinovac Biotech.

Vaccine and Doria become jokes on TV news. *O Governador de São Paulo, João Doria (PSDB), tentou tanto os holofotes que ficou conhecido, internacionalmente. Não da forma positiva, como ele gostaria, mas como motivo de chacota. Em telejornal americano, o casal de apresentadores debochou do resultado dos índices de eficiência da CoronaVac, imunizante da gigante farmacêutica Sinovac em parceria com o Instituto Butantan e Doria foi mencionado como “governador sem credibilidade”. Uma vergonha para todos os brasileiros! (da Redação)*

In January 2021, oxygen shortages in hospitals in Manaus, the capital of Amazonas, featured heavily in the news cycle (<https://www.bbc.com/news/world-latin-america-55670318>, last accessed on 24 January 2022) and similarly resulted in fake news relating to oxygen donations.

Donations of oxygen cylinders. *PRODUÇÃO DA LG PAROU A FABRICA DE AR CONDICIONADO PARA DOAR TODO OXIGÊNIO PARA HOSPITAIS DA REDE PÚBLICA DE MANAUS. A DETERMINAÇÃO FOI DADA PELO PRESIDENTE DA LG*

During February 2021, a time of public frustration regarding the national vaccination timeline plan (<https://edition.cnn.com/2021/02/01/americas/brazil-coronavirus-vaccination/index.html>, last accessed on 24 January 2022), the dominant subject in fake news related to the impact of vaccination on the elderly, suggesting vaccination resulted in amputation or death.

Elderly die after taking the vaccine. *QUATRO idosos da cidade de Canela, NÃO VIRARAM JACARÉ... Se mudaram para a Terra dos Descarnados, horas após serem vacinadas com a VACINA DO BUTANTAN: Terra-Dura-Dura...! Como a cabeça de muita gente séria. Alguém postou: “Não creio pelo fato de terem tomada a vacina da China!”*

Elderly person suffers amputation of three fingers after taking the vaccine. *Chinavirus: Idosa toma vacina e sofre amputação de três dedos devido ao efeito colateral grave*

Fake news often follows the news cycle, which provides it with credibility and garners more attention from the public [24]. In many respects, it leverages the real news cycle for counter-programming. The COVID-19 vaccine was the main focus of fake news that circulated in Brazil during the first year of the pandemic, with themes ranging from the vaccination’s efficacy to its side effects. According to Bivar et al. [95] and Rochel [96], the main reasons that the public believe in and disseminate anti-vaccine messages are based on conspiracy theories surrounding the pharmaceutical industry and governments. These theories promote the idea that information about vaccines, their ingredients and adverse effects is not made publicly available, so that corporations can maximise their profits. Anti-vaccination beliefs are further exacerbated by a general suspicion by members of the public in scientific knowledge and that researchers falsify experiment results regarding the harmful side effects of vaccines in order to present effectiveness [97]. These beliefs gain even greater strength and reach when disseminated on online social networks. Moreover, “social bubbles” (or “echo chambers” as defined by Nguyen [98]) prevent contrary points of view from penetrating such groups, fostering an environment in which their (mis)conceptions are positively reinforced [96].

Previous studies have demonstrated the negative impact of misinformation surrounding COVID-19 on populations’ behaviour during the pandemic. This includes a low adherence to social distancing regulation [15–17] and the strengthening of the anti-vaccine

movement [95]. This analysis reinforces the need for studies analysing characteristics of fake news and fuels the further development of mechanisms to automatically detect and combat the dissemination of such messages.

4.3. Textual Attributes of Fake News

According to Zhou et al. [20], at a semantic level, disinformation-related textual attributes can be grouped along ten dimensions: quantity, complexity, uncertainty, subjectivity, non-immediacy, sentiment, diversity, informality, specificity and readability.

Quantity attributes are related to the amount of information that is present in the message [30]. Figure 5 presents histograms and boxplots relating to the following quantity attributes: (a) No. of characters, (b) No. of words and (c) No. of sentences. In our fake news subset, the mean for No. of characters is 429.40, the mean for No. of words is 87.32 and the mean for No. of sentences is 5.47, as shown in Table 1.

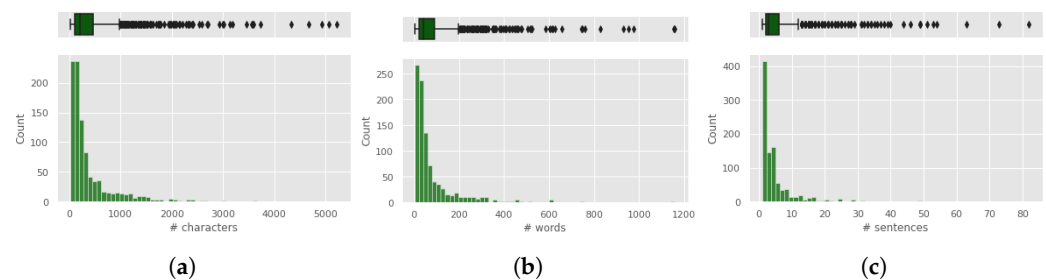


Figure 5. Quantity attributes of fake news items: (a) No. of characters, (b) No. of words and (c) No. of sentences.

Complexity attributes are related to the ratio of characters and words in a message [99,100]. Figure 6 presents histograms and boxplots relating to the following complexity attributes: (a) average No. of characters per word, (b) average No. of words per sentence and (c) average No. of punctuation per sentences. The complexity attributes are similar to those presented in [87,88], in which authors analyse the Fake.Br corpus, a reference corpus focused on fake news in Portuguese. The average No. of characters per word is 5.10 and the average of No. of words per sentence is 18.84.

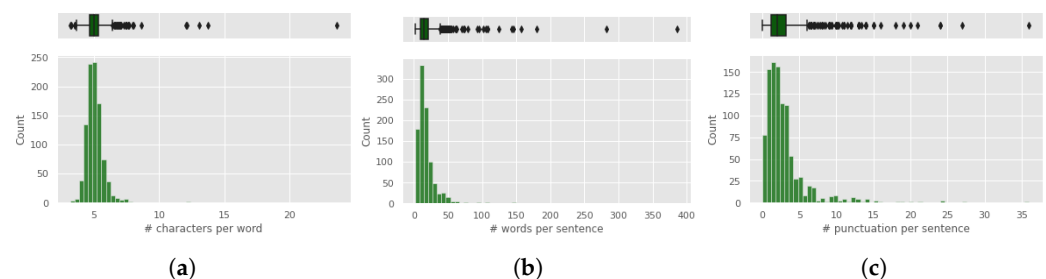


Figure 6. Complexity attributes of fake news items: (a) average No. of characters per word, (b) average No. of words per sentence and (c) average No. of punctuation per sentences.

Mehrabian et al. [101] define non-immediacy as “a measure of the attenuation of directness and intensity of interaction between a communicator and the object of his communication in a verbal message.” Zhou et al. [100] suggest that deceivers tend to disassociate themselves from their deceptive messages in an attempt to reduce accountability and responsibility for their statements, and thus display higher non-immediacy in their language. Figure 7 presents histograms and boxplots relating to the following non-immediacy attributes: (a) No. of self reference, (b) percentage of self reference, (c) No. of group references, (d) percentage of group references, (e) No. of other references and (f) percentage of other references. On average, the amount of other references (1.60) is greater than self reference (0.48) and group reference (0.39), consistent with Zhou et al. [100].

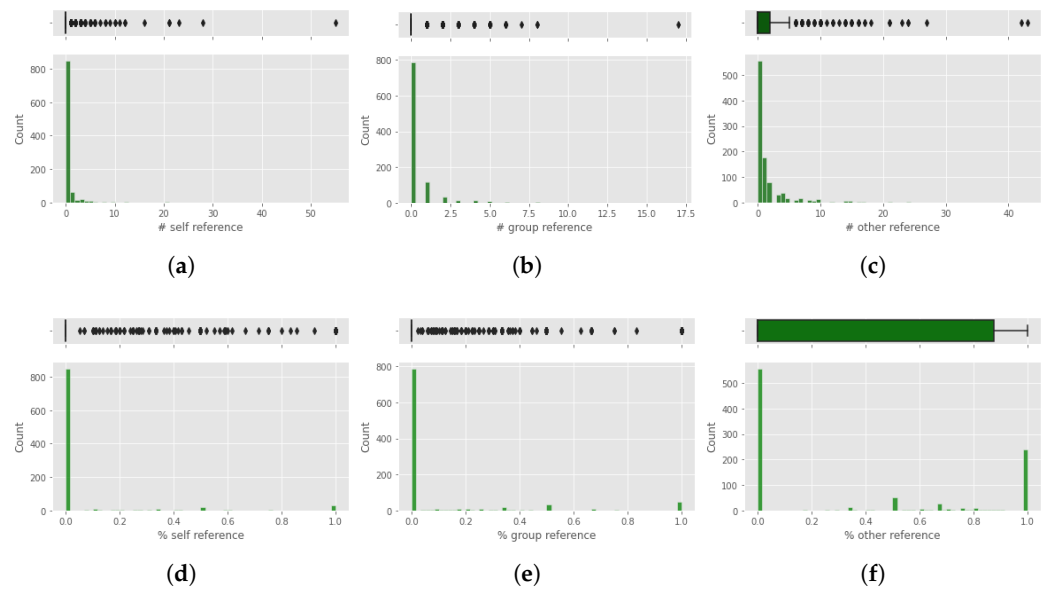


Figure 7. Non-immediacy attributes of fake news items: (a) No. of self reference, (b) percentage of self reference, (c) No. of group references, (d) percentage of group reference, (e) No. of other references and (f) percentage of other references.

Table 1 summarises textural attributes and features for the fake news subset. In general, the fake news subset has the following profile: a message with a few short sentences (75% of content has up to 6 sentences and an average length of 19 words), low complexity (in terms of characters per word, words and punctuation by sentence), low use of numbers (75% has up to five numbers) and exclamation marks and high diversification in terms of words used. This is consistent with previous studies [87,88,100] which have shown that there is a linguistic pattern in fake news, such as a preference for short sentences.

Table 1. Textual attributes and features of fake news that circulated in Brazil from January 2020 to February 2021.

Attribute Type	Feature	Mean	Std	Min	Q3	Max
Quantity	No. of characters	429.40	623.24	22	447.50	5221
	No. of words	87.32	130.38	4	91	1159
	No. of sentences	5.47	7.93	1	6	82
Complexity	Average No. of characters per word	5.10	1.04	3.03	5.34	23.63
	Average No. of words per sentence	18.84	21.48	2	20.91	386
	Average No. of punctuations per sentence	2.88	3.06	0	3.27	36
Uncertainty	No. of numbers	5.81	14.76	0	5	221
	Percentage of numbers	0.08	0.18	0	0.08	2.66
	No. of question markers	0.60	1.91	0	0	27
	Percentage of question markers	0.04	0.14	0	0	1
Non-immediacy	No. of self reference	0.48	2.57	0	0	55
	Percentage of self reference	0.06	0.20	0	0	1
	No. of group reference	0.39	1.14	0	0	17
	Percentage of group reference	0.10	0.25	0	0	1
	No. of other reference	1.60	3.72	0	1	43
	Percentage of other reference	0.34	0.43	0	0.84	1
Sentiment	No. of exclamation marks	1.13	2.41	0	1	24
	Percentage of exclamation marks	0.11	0.21	0	0.12	1
Diversity	No. of unique words	61.84	72.25	4	70	565
	Percentage of unique words	0.85	0.12	0.26	0.95	1

5. Detecting Fake News Using Machine Learning and Deep Learning

As discussed in Section 3, machine learning and deep learning models have been widely applied in NLP. In this paper, we compare the performance of four supervised machine learning techniques as per [59]—support vector machine (SVM), Random Forest, gradient boosting and naive Bayes—against four deep learning models—LSTM, Bi-LSTM, GRU and Bi-GRU as per [102].

SVM is a non-parametric technique [103] capable of performing data classification; it is commonly used in images and texts [104]. SVM is based on hyper-plane construction and its goal is to find the optimal separating hyper-plane where the separating margin between two classes should be maximised. Random Forest and gradient boosting are tree-based classifiers that perform well for text classification [104,105]. Random Forest unifies several decision trees, building them randomly from a set of possible trees with K random characteristics in each node. By using an ensemble, gradient boosting increases the robustness of classifiers while decreasing their variances and biases [105]. Naive Bayes is a traditional classifier based on the Bayes theorem. Due to its relatively low memory use, it is considered computationally inexpensive compared with other approaches [104]. Naive Bayes assigns the most likely class to a given example described by its characteristic vector. In this study, we consider the naive Bayes algorithm for multinomially distributed data.

As discussed, we evaluate four deep learning models—two unidirectional RNNs (LSTM and GRU) and two bidirectional RNNs (Bi-LSTM and Bi-GRU). Unlike unidirectional RNNs that process the input in a sequential manner and ignoring future context, in bidirectional RNNs (Bi-RNNs) the input is presented forwards and backwards to two separate recurrent networks, and both are connected to the same output layer [106]. In this work, we use two types of bi-RNN: bidirectional long short-term memory (Bi-LSTM) and bidirectional gated recurrent unit (Bi-GRU) as per [102], as they demonstrated a good performance in text classification tasks.

The machine learning and deep learning models were implemented on Python using the following libraries: Tensorflow (<https://www.tensorflow.org/?hl=pt-br>, last accessed on 24 January 2022), Scikit-learn (<https://scikit-learn.org/>, last accessed on 24 January 2022), and keras (<https://keras.io/>, last accessed on 24 January 2022).

To train and test the models, we use the dataset presented in Section 4.1 containing 1047 fake news items and 10,285 true news items. We applied a random undersampling technique to balance the dataset, where the largest class is randomly trimmed until it is the same size as the smallest class. The final dataset used to train and test the models comprises 1047 fake news items and 1047 true news items, totalling 2094 items. A total of 80% of the dataset was allocated for training and 20% for testing.

5.1. Evaluation Metrics

To evaluate the performance of the models, we consider the following metrics as per [59]: accuracy, precision, recall, specificity and F1 score. These metrics are based on a confusion matrix, a cross table that records the number of occurrences between the true classification and the classification predicted by the model [107]. It is composed of true positive (TP), true negative (TN), false positive (FP) and false negative (FN).

Accuracy is the percentage of correctly classified instances over the total number of instances [108]. It is calculated as the sum of TP and TN divided by the total of samples, as shown in Equation (1).

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Precision is the number of class members classified correctly over the total number of instances classified as class members [108]. It is calculated as the number of TP divided by the sum of TP and FP, as shown in Equation (2).

$$precision = \frac{TP}{TP + FP} \quad (2)$$

Recall (sensitivity or true positive rate) is the number of class members classified correctly over the total number of class members [108]. It is calculated as the number of TP divided by the sum of TP and FN, as shown in Equation (3).

$$recall = \frac{TP}{TP + FN} \quad (3)$$

Specificity (or true negative rate) is the number of class members classified correctly as negative. It is calculated by the number of TN divided by the sum of TN and FP, as per Equation (4).

$$specificity = \frac{TN}{TN + FP} \quad (4)$$

To address vulnerabilities in machine learning systems designed to optimise precision and recall, weighted harmonic means can be used to balance between these metrics [109]. This is known as the F1 score. It is calculated as per Equation (5).

$$F1\text{-score} = 2 \times \frac{precision \times recall}{precision + recall} \quad (5)$$

5.2. Experiments

In their evaluation of machine learning techniques for junk e-mail spam detection, Méndez et al. [110] analysed the impact of stop-word removal, stemming and different tokenization schemes on the classification task. They argued that “spammers often introduce “noise” in their messages using phrases like “MONEY!!”, “FREE!!!” or placing special characters into the words like “R-o-l-e?x””. As fake news can be viewed as similar type of noise within a document, we define two experiments in order to evaluate the impact of removing such noise from news items: (a) applying text preprocessing techniques and (b) using the raw text of the fake news item.

For the first experiment, we remove the stop words, convert text to lower case and apply vectorization to convert text to a matrix of token counts. For the second experiment, we use the raw text without removing stop words and without converting text to lowercase. We also apply vectorization to convert text to a matrix of token counts, similar to the first experiment.

Specifically for deep learning models, we used an embedding layer built using the FastText library (<https://fasttext.cc/>, last accessed on 1 March 2022), which has a pre-trained word vector available in Portuguese. Padding was also performed so that all inputs had the same size. All deep learning models were trained using 20 epochs with a dropout in the recurrent layers with a probability of 20% to mitigate against overfitting [111]. Grid search was used to optimise model hyper-parameters [112].

It is hard to make a direct comparison with extant studies due to challenges in reproducibility. For example, similar data may not be available. Notwithstanding this, to facilitate comparison we benchmarked against the study reported by Paixão et al. [113]. They also propose classifiers for fake news detection although without a focus on COVID-19; they use fake news relating to politics, TV shows, daily news, technology, economy and religion. We replicated one of their deep learning models, a CNN (see Table 2), and tested it with our dataset for comparison.

Table 2. Configuration of the CNN model proposed by Paixão et al. [113].

Layers	Shape
Embedding	(200, 300)
Conv1d	(198, 100)
Pooling1d	(66, 100)
Flatten	(6600)
Dense	(1)

5.3. Evaluation

5.3.1. Experiment 1—Applying Text Preprocessing

Tables 3 and 4 present the parameters and levels used in the grid search for machine learning and deep learning models, respectively, when applying text preprocessing techniques (Experiment 1). The values in bold indicate the best configuration of each model based on the F1 score.

Table 3. Grid search parameters and levels for machine learning models when applying text preprocessing techniques.

Models	Parameter	Levels
SVM	Kernel type	linear, rbf , sigmoid
	Regularisation parameter	1.0, 2.0, 3.0, 4.0
Random Forest	Number of trees in the forest	100, 200 , 300, 400
	Number of features	sqrt, log2
Gradient Boosting	Number of boosting stages	100, 500, 1000
	Learning rate	0.1 , 0.2, 0.3, 0.4
Naive Bayes	alpha	1.0 , 5.0, 10.0

Table 4. Grid search parameters and levels for deep learning models when applying text preprocessing techniques.

Models	Parameter	Levels
LSTM	Number of layers	1, 2, 3
	Number of nodes	50 , 100, 200
GRU	Number of layers	1, 2, 3
	Number of nodes	50, 100, 200
bi-LSTM	Number of layers	1, 2, 3
	Number of nodes	50, 100, 200
bi-GRU	Number of layers	1, 2, 3
	Number of nodes	50, 100 , 200

Table 5 presents the classification results of ML and deep learning models using the configuration selected by the grid search technique (see Tables 3 and 4). The test experiments were executed ten times, the metric results are the mean and their respective standard deviation.

For the machine learning models, the best accuracy (92.36%), precision (96.31%) and F1 score (92.88%) were achieved by the SVM model. The best recall (92.57%) was achieved by the naive Bayes, and the best specificity (95.80%) was achieved by the Random Forest. Gradient boosting presented the worst machine learning performance. In general, the machine learning models achieved recall at levels greater than 81.33%, the recall obtained by the gradient boosting model.

Table 5. Evaluation of proposed models regarding accuracy, precision, recall and F1 score (in % with its respective standard deviation) when applying text preprocessing techniques. Please note that the Random Forest and naive Bayes models presented a very low standard deviation, and thus it is represented as zero.

Models	Accuracy	Precision	Recall	Specificity	F1 Score
Machine learning					
SVM	92.36 (± 0.000)	96.31 (± 0.000)	89.69 (± 0.000)	95.69 (± 0.000)	92.88 (± 0.000)
Random Forest	91.05 (± 0.003)	96.31 (± 0.005)	87.25 (± 0.008)	95.80 (± 0.006)	91.55 (± 0.003)
Gradient Boosting	87.06 (± 0.005)	90.87 (± 0.003)	81.33 (± 0.009)	92.39 (± 0.003)	85.83 (± 0.006)
Naive Bayes	87.82 (± 0.000)	83.95 (± 0.000)	92.57 (± 0.000)	83.41 (± 0.000)	88.00 (± 0.000)
Deep learning					
LSTM	88.87 (± 0.014)	93.68 (± 0.026)	83.61 (± 0.043)	94.16 (± 0.030)	88.24 (± 0.018)
GRU	88.80 (± 0.013)	91.80 (± 0.024)	86.04 (± 0.047)	91.71 (± 0.031)	88.69 (± 0.016)
bi-LSTM	89.42 (± 0.009)	90.45 (± 0.016)	86.34 (± 0.026)	92.08 (± 0.016)	88.30 (± 0.012)
bi-GRU	88.78 (± 0.010)	88.73 (± 0.017)	86.85 (± 0.022)	90.44 (± 0.018)	87.74 (± 0.012)
CNN [113]	91.22 (± 0.003)	93.19 (± 0.009)	88.54 (± 0.007)	93.79 (± 0.009)	90.80 (± 0.003)

There was significantly more variation in the deep learning models. The CNN model proposed by Paixão et al. [113] outperformed our proposed models in three metrics: accuracy (91.22%), recall (88.54%) and F1 score (90.80%). This is explained by the relationships between the the F1 score and the precision and recall metrics, i.e., the F1 score is the harmonic mean between the precision and recall. Paixão et al.'s CNN model obtained high levels of those metrics. Our LSTM model presented the best precision (93.68%) and the best specificity (94.16%). These metrics are very important in the context of fake news detection. Specificity assesses the method's ability to detect true negatives, that is, correctly detect true news as in fact true. On the other hand, precision assesses the ability to classify fake news as actually false. Having a model that achieves good results in these metrics is of paramount importance for automatic fake news classification. It is also important to mention that the bidirectional models (Bi-LSTM and Bi-GRU) presented competitive results with other models but did not outperform any other unidirectional model.

Figure 8 shows the receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC) of all models when applying text preprocessing techniques. Gradient boosting presented the worst result with an AUC of 0.8552, which is expected since it presented the worst recall result, as shown in Table 5.

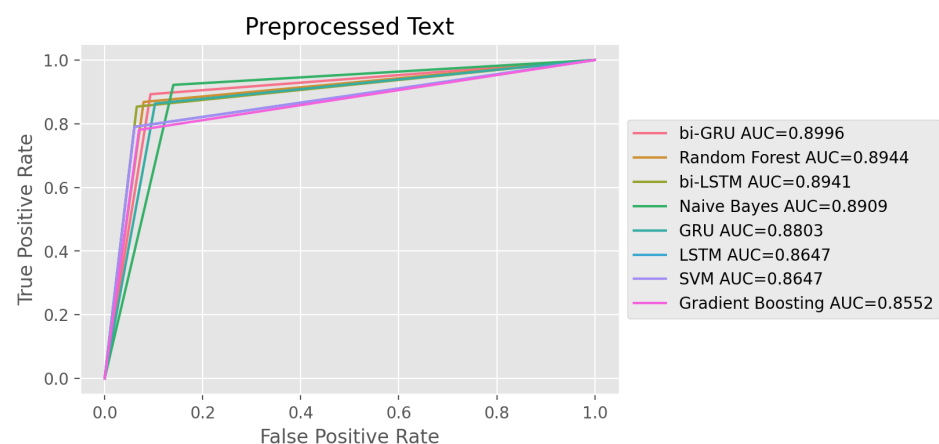


Figure 8. ROC curve and AUC of proposed models when applying text preprocessing techniques.

Although the naive Bayes presented the best recall, it also presented the highest false positive rate resulting in an AUC of 0.8909. The Bi-GRU model presented the highest AUC value (0.8996), followed by Random Forest (0.8944) and Bi-LTSM (0.8941). It is important to note that the bidirectional models outperformed the respective unidirectional models in

obtaining the best AUC results. The Bi-GRU and Bi-LSTM presented AUC values of 0.8996 and 0.8941, respectively, while the GRU and LSTM presented AUC values of 0.8803 and 0.8647, respectively.

For illustration purposes, we selected some fake news that were misclassified. We added the complete text of news, but the models process the sentences in lowercase and without stop words. “Hospital de campanha do anhembi vazio sem colchões para os pacientes e sem atendimento” was misclassified by the SVM model and “Agora é obrigatório o Microchip na Austrália! Muito protesto por lá.. Australianos vão às ruas para protestar contra o APP de rastreamento criado para controlar pessoas durante o Covid-19, contra exageros nas medidas de Lockdown, distanciamento social, uso de máscaras, Internet 5G, vacina do Bill Gates e microchip obrigatório. Dois líderes do protesto foram presos e multados.” was misclassified by the LSTM model. In addition, we noted that the fake news misclassified by the SVM model was also misclassified by the LSTM model. Consistent with Cabral et al. [91] and Martins et al. [92], misclassification would seem to be more likely where text is too short to be classified correctly by the models.

5.3.2. Experiment 2: Using Raw Text

Tables 6 and 7 present the parameters and levels used in the grid search for machine learning and deep learning models, respectively, when we use raw text of the news items (Experiment 2). The parameters and levels are the same as those used in Experiment 1. Again, the values in bold indicate the best configuration of each model based on F1 score.

Table 6. Grid search parameters and levels for machine learning models when using raw text of the messages (Experiment 2).

Models	Parameter	Levels
SVM	Kernel type	linear, rbf , sigmoid
	Regularisation parameter	1.0, 2.0, 3.0, 4.0
Random Forest	Number of trees in the forest	100, 200, 300, 400
	Number of features	sqrt, log2
Gradient Boosting	Number of boosting stages	100, 500, 1000
	Learning rate	0.1, 0.2, 0.3 , 0.4
Naive Bayes	alpha	1.0 , 5.0, 10.0

Table 7. Grid search parameters and levels for deep learning models when using raw text of the messages.

Models	Parameter	Levels
LSTM	Number of layers	1, 2 , 3
	Number of nodes	50 , 100, 200
GRU	Number of layers	1, 2 , 3
	Number of nodes	50 , 100, 200
bi-LSTM	Number of layers	1, 2 , 3
	Number of nodes	50 , 100, 200
bi-GRU	Number of layers	1, 2 , 3
	Number of nodes	50 , 100, 200

Table 8 presents the classification results of machine learning and deep learning models when using raw text of the messages and the configuration selected by the grid search technique (see Tables 6 and 7). The test experiments were executed 10 times; the metric results are the mean and their respective standard deviation.

Table 8. Evaluation of proposed models regarding accuracy, precision, recall and F1 score (in % with its respective standard deviation) using raw text of the messages. Please note that Random Forest and naive Bayes presented a very low standard deviation and it is represented as zero.

Models	Accuracy	Precision	Recall	Specificity	F1 score
Machine learning					
SVM	93.79 (± 0.000)	96.23 (± 0.000)	90.40 (± 0.000)	96.83 (± 0.000)	93.22 (± 0.000)
Random Forest	94.36 (± 0.003)	97.91 (± 0.005)	90.00 (± 0.005)	98.28 (± 0.004)	93.78 (± 0.003)
Gradient Boosting	93.36 (± 0.006)	95.85 (± 0.004)	89.84 (± 0.011)	96.51 (± 0.003)	92.74 (± 0.007)
Naive Bayes	88.78 (± 0.000)	84.79 (± 0.000)	92.92 (± 0.000)	85.06 (± 0.000)	88.67 (± 0.000)
Deep learning					
LSTM	92.98 (± 0.007)	93.66 (± 0.022)	91.46 (± 0.031)	94.34 (± 0.022)	92.47 (± 0.009)
GRU	93.26 (± 0.008)	93.17 (± 0.022)	92.67 (± 0.033)	93.80 (± 0.022)	92.84 (± 0.010)
bi-LSTM	94.34 (± 0.007)	95.39 (± 0.011)	92.52 (± 0.020)	95.97 (± 0.011)	93.91 (± 0.008)
bi-GRU	94.41 (± 0.006)	95.01 (± 0.017)	93.13 (± 0.016)	95.56 (± 0.017)	94.03 (± 0.006)
CNN [113]	93.15 (± 0.005)	98.20 (± 0.009)	87.61 (± 0.009)	98.46 (± 0.004)	92.56 (± 0.005)

For the machine learning models, Random Forest performed best across four metrics: accuracy (94.36%), precision (97.91%), specificity (98.28%) and F1 score (93.78%). The best recall (92.92%) was achieved by the naive Bayes. In contrast with Experiment 1, the SVM did not perform the best in any metric, although its metrics were all above 90% and close to the Random Forest values. Similar to Experiment 1, gradient boosting performed the worst of the machine learning models evaluated.

Regarding the evaluation of deep learning models and in contrast with Experiment 1, the bidirectional models presented better results than their respective unidirectional models. The Bi-LSTM presented the best accuracy (94.34%) while the Bi-GRU presented the best levels of recall (93.13%) and F1 score (94.03%). The CNN model proposed by Paixão et al. [113] obtained the best results for precision (98.20%) and specificity (98.46%), however, it presented lower recall values (87.61%) which also impacts the F1 score.

Figure 9 presents the ROC curve and AUC results for the models when using raw text. The results are slightly closer to those presented in Figure 8, showing that the proposed models are able to obtain good results using raw text or preprocessed texts. However, in contrast to using preprocessing text techniques, the Random Forest outperformed all other models, since it presented the lowest false positive rate and a relatively good true positive rate. The second and third best performing models were, in order, the Bi-LSTM and Bi-GRU models with AUC values of 0.8948 and 0.8927, respectively. The LSTM and SVM models presented the same AUC result (0.8924), while the GRU and naive Bayes models had lower results than those shown in Figure 8, with AUCs of 0.8909 and 0.8729, respectively. The gradient boosting model presented the worst results for AUC, 0.843, similar to applying preprocessing techniques.

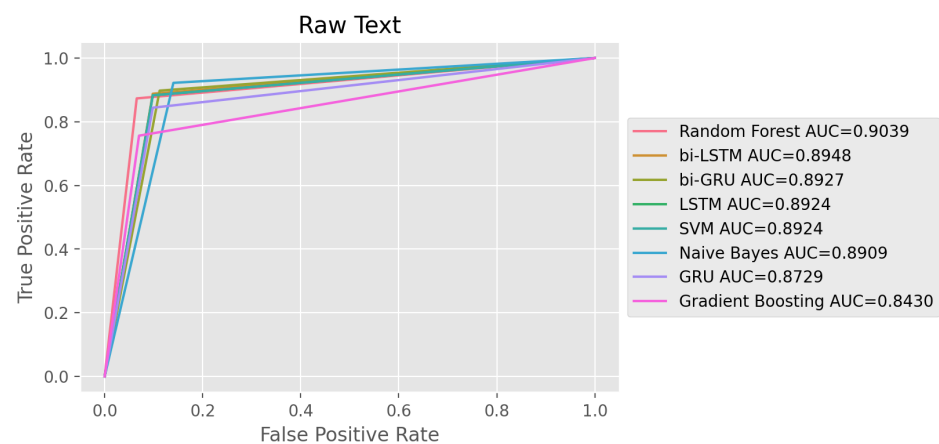


Figure 9. ROC curve and AUC of proposed models when using raw text.

Again, for illustration purposes, we selected samples of fake news that were misclassified by the models. “*Ative sua conta grátis pelo PERÍODO DE ISOLAMENTO! NETFLIX-USA.NET Netflix Grátis contra COVID-19 Ative sua conta grátis pelo PERÍODO DE ISOLAMENTO!*” was misclassified by the Random Forest model. It is interesting to note that this fake news is not about the direct impact of COVID-19, but about an indirect impact, i.e., the availability of free Netflix during the pandemic. “*Não há como ser Ministro da Saúde num país onde o presidente coloca sua mediocridade e ignorância a frente da ciência e não sente pelas vítimas desta pandemia. Nelson Teich. URGENTE Nelson Teich pede exoneração do Ministério da Saúde por não aceitar imposições de um Presidente medíocre e ignorante.*” is a message that was misclassified by the Bi-GRU model. Again, this is not directly about the pandemic, but rather the resignation of Brazil’s Minister of Health during the pandemic. In contrast to Experiment 1 misclassification, here we believe the news item is too long, thus compromising the model’s ability to classify it correctly.

5.4. Discussion, Challenges and Limitations

One would normally expect deep learning models to outperform machine learning models due their ability to deal with more complex multidimensional problems. In our study, when applying text preprocessing techniques (Table 5), machine learning presented better results than deep learning across all metrics analysed. In contrast to previous works that used the Fake.Br corpus [87–90], a dataset composed of 7200 messages, we used an equally balanced dataset composed of only 2094 messages, which could explain the relatively poor performance of the deep learning models. Deep learning models require large volumes of data for training [114]. Since the dataset used was relatively small, we posit that the deep learning models were not able to adequately learn the fake news patterns from the data set once stop words were removed. The creation of a sufficiently large dataset is not an insignificant challenge in the context of COVID-19 fake news in Portuguese.

In general, the best results were obtained when using the raw text of the news items, i.e., without removing stop words and converting text to lowercase. Random Forest presented the best results for accuracy (94.36%), precision (97.91%) and specificity (98.28%); the Bi-GRU model presented the best recall (93.13%) and F1 score (94.03%). The usage of raw text had a significant impact on the performance of the deep learning models performance; all deep learning metrics in Experiment 2 were above 91% (Table 8). The performance of the machine learning models was also impacted positively. This is consistent with Méndez et al. [110] suggestion that that noise removal may hamper the classification of non-traditional documents, such as junk e-mail, and in this case, fake news. In effect, pre-processing made the fake news item structure closer to the true news item structure, making it difficult to classify them. This is supported by the improvement in the recall metric in both machine learning and deep learning models. In our context, the recall assesses the model’s performance in correctly classifying fake news. A considerable improvement was achieved in practically all models, except naive Bayes, which had already achieved a good result with the pre-processed base (92.57%). Considering the pre-processed base, the recall of most models were below 90% (disregarding the naive Bayes), where the best recall of the machine and deep learning models were, respectively, SVM (89.69%) and Bi-GRU (86.85%). When using raw text, most models had a recall above 90%, with the exception of gradient boosting which had the worst result (89.84%). By running both experiments, one can see the impact of pre-processing clearly—keeping the stop words and the capital letters in the text contributes to improved correct classification of fake news.

This study has a number of limitations. A range of parameters and levels used in the grid search for finding the best configuration for each model were identified. Other configurations may result in better results. As discussed earlier, larger datasets prepared by specialist fact-checking organisations may result in different results. The evaluation of machine learning and deep learning models using accuracy, precision, recall, specificity and F1score presents challenges in describing differences in results. Complementarity has been suggested as a potential solution, particularly where F1 scores are very close [109].

6. Conclusions and Future Work

In February 2020, two months after the disclosure of the first COVID-19 case in Wuhan, China, the WHO published a report declaring that the pandemic was accompanied by an “infodemic”. An exceptional volume of information about the COVID-19 outbreak started to be produced, however, not always from reliable sources, making it difficult to source adequate guidance on this new health threat. The spread of fake news represents a serious public health issue that can endanger lives around the world.

While machine learning does not usually require deep linguistic knowledge, pluricentric and highly diglossic languages, such as Brazilian Portuguese, require specific solutions to achieve accurate translation and classification, particularly where meaning may be ambiguous or misleading as in the case of fake news. Such challenges are further exacerbated when new terms evolve or are introduced in to linguistic repertoires, as in the case of COVID-19. Given the scale and negative externalities of health misinformation due to COVID-19, it is not surprising that there has been an increase in recent efforts to produce fake news datasets in Brazilian Portuguese. We add to this effort by providing a new dataset composed of 1047 fake news items and 10,285 news items related to COVID-19 in Portuguese that circulated in Brazil in the last year. Based on the fake news items, we performed an exploratory data analysis, exploring the main themes and their textual attributes. Vaccines and vaccination were the central themes of the fake news that circulated in Brazil during the focal period.

We also proposed and evaluated machine learning and deep learning models to automatically classify COVID-19 fake news in Portuguese. The results show that, in general, SVM presented the best results, achieving 98.28% specificity. All metrics of all models were improved when using the raw text of the messages (the only exception was the precision of SVM, but the difference was very little, 0.08%).

As future work, we plan to increase the dataset with more fake news featuring specific L-variant and H-variant classifications as well as other dialectal features, as well as the source of the fake news (websites, WhatsApp, Facebook, Twitter etc.), and the type of multimedia used (video, audio and image). This extended dataset will not only aid in the refinement of fake news detection but allow researchers to explore other aspects of the fake news phenomenon during COVID-19, including detection of pro-vaccination, anti-vaccination and vaccine-hesitant users, content, motivation, topics and targets of content and other features including stigma. Furthermore, future research should consider the diffusion and impact of fake news and explore both the extent of propagation, the type of engagement and actors involved, including the use of bots. The detection of new types of fake news, particularly in a public health context, can inform public health responses but also optimise platform moderation systems. To this end, research on the use of new transformer-based deep learning architectures such as BERT and GPT-3 may prove fruitful.

Author Contributions: Conceptualization, P.T.E., G.R.N.C. and T.L.; methodology, P.T.E. and T.L.; software, P.T.E., G.L.S., M.E.d.L.X. and G.R.N.C.; validation, P.T.E., L.C.d.L., I.S. and T.L.; formal analysis, P.T.E., L.C.d.L., I.S., A.E. and T.L.; investigation, P.T.E., L.C.d.L., I.S. and T.L.; resources, P.T.E., M.E.d.L.X. and G.R.N.C.; data curation, P.T.E., G.L.S., M.E.d.L.X. and G.R.N.C.; writing—original draft preparation, P.T.E., M.E.d.L.X. and G.R.N.C.; writing—review and editing, P.T.E., L.C.d.L., I.S., A.E. and T.L.; visualization, P.T.E.; supervision, P.T.E. and T.L.; project administration, P.T.E.; funding acquisition, T.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Irish Institute of Digital Business (IIDB).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: <https://github.com/rhuancampos/covid-19-rumor> (accessed on 24 January 2022).

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

References

1. Bujnowska-Fedak, M.M.; Waligóra, J.; Mastalerz-Migas, A. The internet as a source of health information and services. In *Advancements and Innovations in Health Sciences*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 1–16.
2. Ofcom. Online Nation 2020 Report. 2020. Available online: https://www.ofcom.org.uk/__data/assets/pdf_file/0027/196407/online-nation-2020-report.pdf (accessed on 24 January 2022).
3. Eurobarometer Flash. European citizens' digital health literacy. *Rep. Eur. Comm.* **2014**. [CrossRef]
4. Lynn, T.; Rosati, P.; Leoni Santos, G.; Endo, P.T. Sorting the Healthy Diet Signal from the Social Media Expert Noise: Preliminary Evidence from the Healthy Diet Discourse on Twitter. *Int. J. Environ. Res. Public Health* **2020**, *17*, 8557. [CrossRef] [PubMed]
5. Sinapuelas, I.C.; Ho, F.N. Information exchange in social networks for health care. *J. Consum. Mark.* **2019**, *36*, 692–702. [CrossRef]
6. Allem, J.P.; Ferrara, E. Could social bots pose a threat to public health? *Am. J. Public Health* **2018**, *108*, 1005. [CrossRef]
7. Broniatowski, D.A.; Jamison, A.M.; Qi, S.; AlKulaib, L.; Chen, T.; Benton, A.; Quinn, S.C.; Dredze, M. Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. *Am. J. Public Health* **2018**, *108*, 1378–1384. [CrossRef] [PubMed]
8. van der Linden, S.; Roozenbeek, J.; Compton, J. Inoculating against Fake News about COVID-19. *Front. Psychol.* **2020**, *11*, 2928. [CrossRef] [PubMed]
9. Evanega, S.; Lynas, M.; Adams, J.; Smolenyak, K.; Insights, C.G. Coronavirus Misinformation: Quantifying Sources and Themes in the COVID-19 'Infodemic'. *Jmir Prepr.* 2020. Available online: <https://allianceforscience.cornell.edu/wp-content/uploads/2020/09/Evanega-et-al-Coronavirus-misinformationFINAL.pdf> (accessed on 24 January 2022).
10. WHO. Munich Security Conference. 2020. Available online: <https://www.who.int/director-general/speeches/detail/munich-security-conference> (accessed on 9 January 2021).
11. Organisation, W.H. Coronavirus Disease (COVID-19) Advice for the Public: Mythbusters. 2020. Available online: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters> (accessed on 24 January 2022).
12. WHO. COVID-19 in Brazil. 2020. Available online: <https://covid19.who.int/region/amro/country/br> (accessed on 9 January 2021).
13. Endo, P.T.; Silva, I.; Lima, L.; Bezerra, L.; Gomes, R.; Ribeiro-Dantas, M.; Alves, G.; Monteiro, K.H.d.C.; Lynn, T.; Sampaio, V.d.S. # StayHome: Monitoring and benchmarking social isolation trends in Caruaru and the Região Metropolitana do Recife during the COVID-19 pandemic. *Rev. Soc. Bras. Med. Trop.* **2020**, *53*. [CrossRef]
14. Anderson, R.M.; Heesterbeek, H.; Klinkenberg, D.; Hollingsworth, T.D. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *Lancet* **2020**, *395*, 931–934. [CrossRef]
15. Ajzenman, N.; Cavalcanti, T.; Da Mata, D. More than Words: Leaders' Speech and Risky Behavior during a Pandemic. 2020. Available online: <https://ssrn.com/abstract=3582908> (accessed on 24 January 2022).
16. Gugushvili, A.; Koltai, J.; Stuckler, D.; McKee, M. Votes, populism, and pandemics. *Int. J. Public Health* **2020**, *65*, 721–722. [CrossRef]
17. Mariani, L.; Gagete-Miranda, J.; Retti, P. Words can hurt: How political communication can change the pace of an epidemic. *Covid Econ.* **2020**, *12*, 104–137.
18. Cardoso, C.R.d.B.; Fernandes, A.P.M.; Santos, I.K.F.d.M. What happens in Brazil? A pandemic of misinformation that culminates in an endless disease burden. *Rev. Soc. Bras. Med. Trop.* **2021**, *54*. Available online: <https://www.scielo.br/j/rsbmt/a/x6z3v5bHDCKPvbdFD7CvY3f/?lang=en> (accessed on 24 January 2022). [CrossRef] [PubMed]
19. de Figueiredo, A.; Simas, C.; Karafillakis, E.; Paterson, P.; Larson, H.J. Mapping global trends in vaccine confidence and investigating barriers to vaccine uptake: A large-scale retrospective temporal modelling study. *Lancet* **2020**, *396*, 898–908. [CrossRef]
20. Zhou, X.; Zafarani, R. A survey of fake news: Fundamental theories, detection methods, and opportunities. *Acm Comput. Surv. (CSUR)* **2020**, *53*, 1–40. [CrossRef]
21. da Saúde Governo do Brasil, M. Novo Coronavírus Fake News. 2020. Available online: <https://www.saude.gov.br/component/tags/tag/novo-coronavirus-fake-news> (accessed on 27 July 2020).
22. Ghenai, A.; Mejova, Y. Catching Zika fever: Application of crowdsourcing and machine learning for tracking health misinformation on Twitter. *arXiv* **2017**, arXiv:1707.03778.
23. Swire-Thompson, B.; Lazer, D. Public health and online misinformation: Challenges and recommendations. *Annu. Rev. Public Health* **2020**, *41*, 433–451. [CrossRef]
24. Zhang, X.; Ghorbani, A.A. An overview of online fake news: Characterization, detection, and discussion. *Inf. Process. Manag.* **2020**, *57*, 102025. [CrossRef]
25. Silva, R.V.M. *Ensaio Para Uma Sócio-História do Português Brasileiro*; Parábola Editorial: São Paulo, Brazil, 2004.
26. da Silva, A. Measuring and parameterizing lexical convergence and divergence between European and Brazilian Portuguese. *Adv. Cogn. Socioling.* **2010**, *45*, 41.
27. Piller, I.; Zhang, J.; Li, J. Linguistic diversity in a time of crisis: Language challenges of the COVID-19 pandemic. *Multilingua* **2020**, *39*, 503–515. [CrossRef]
28. Zhou, X.; Zafarani, R. Fake news: A survey of research, detection methods, and opportunities. *arXiv* **2018**, arXiv:1812.00315.
29. Undeutsch, U. Beurteilung der Glaubhaftigkeit von Aussagen (Evaluation of statement credibility). In *Handbuch der Psychologie, Vol. 11: Forensische Psychologie*; Undeutsch, U., Ed.; Hogrefe: Göttingen, Germany, 1967; pp. 26–181.

30. McCornack, S.A.; Morrison, K.; Paik, J.E.; Wisner, A.M.; Zhu, X. Information manipulation theory 2: A propositional theory of deceptive discourse production. *J. Lang. Soc. Psychol.* **2014**, *33*, 348–377. [[CrossRef](#)]
31. Zuckerman, M.; DePaulo, B.M.; Rosenthal, R. Verbal and nonverbal communication of deception. In *Advances in Experimental Social Psychology*; Elsevier: Amsterdam, The Netherlands, 1981; Volume 14, pp. 1–59.
32. Castelo, S.; Almeida, T.; Elghafari, A.; Santos, A.; Pham, K.; Nakamura, E.; Freire, J. A topic-agnostic approach for identifying fake news pages. In Proceedings of the Companion 2019 World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019; pp. 975–980.
33. Moorhead, S.A.; Hazlett, D.E.; Harrison, L.; Carroll, J.K.; Irwin, A.; Hoving, C. A new dimension of health care: Systematic review of the uses, benefits, and limitations of social media for health communication. *J. Med. Internet Res.* **2013**, *15*, e85. [[CrossRef](#)] [[PubMed](#)]
34. Chew, C.; Eysenbach, G. Pandemics in the age of Twitter: Content analysis of Tweets during the 2009 H1N1 outbreak. *PLoS ONE* **2010**, *5*, e14118. [[CrossRef](#)] [[PubMed](#)]
35. Kovic, I.; Lulic, I.; Brumini, G. Examining the medical blogosphere: An online survey of medical bloggers. *J. Med. Internet Res.* **2008**, *10*, e28. [[CrossRef](#)]
36. Zhao, Y.; Zhang, J. Consumer health information seeking in social media: A literature review. *Health Inf. Libr. J.* **2017**, *34*, 268–283. [[CrossRef](#)] [[PubMed](#)]
37. Mavragani, A.; Ochoa, G. The internet and the anti-vaccine movement: Tracking the 2017 EU measles outbreak. *Big Data Cogn. Comput.* **2018**, *2*, 2. [[CrossRef](#)]
38. Hughes, B.; Joshi, I.; Wareham, J. Health 2.0 and Medicine 2.0: Tensions and controversies in the field. *J. Med. Internet Res.* **2008**, *10*, e23. [[CrossRef](#)]
39. Van De Belt, T.H.; Engelen, L.J.; Berben, S.A.; Schoonhoven, L. Definition of Health 2.0 and Medicine 2.0: A systematic review. *J. Med. Internet Res.* **2010**, *12*, e18. [[CrossRef](#)]
40. Pagoto, S.; Waring, M.E.; Xu, R. A Call for a Public Health Agenda for Social Media Research. *J. Med. Internet Res.* **2019**, *21*, e16661. [[CrossRef](#)]
41. Bridgman, A.; Merkley, E.; Loewen, P.J.; Owen, T.; Ruths, D.; Teichmann, L.; Zhilin, O. The causes and consequences of COVID-19 misperceptions: Understanding the role of news and social media. *Harv. Kennedy Sch. Misinform. Rev.* **2020**, *1*. Available online: <https://misinformreview.hks.harvard.edu/article/the-causes-and-consequences-of-covid-19-misperceptions-understanding-the-role-of-news-and-social-media/> (accessed on 24 January 2022). [[CrossRef](#)]
42. Matos, C. *Journalism and Political Democracy in Brazil*; Lexington Books: Lanham, MD, USA, 2008.
43. Milhorange, F.; Singer, J. Media trust and use among urban news consumers in Brazil. *Ethical Space Int. J. Commun. Ethics* **2018**, *15*, 56–65.
44. Reuters Institute. Digital News Report 2020. Available online: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-06/DNR_2020_FINAL.pdf (accessed on 10 January 2021).
45. Ricard, J.; Medeiros, J. Using misinformation as a political weapon: COVID-19 and Bolsonaro in Brazil. *Harv. Kennedy Sch. Misinf. Rev.* **2020**. Available online: <https://misinformreview.hks.harvard.edu/article/using-misinformation-as-a-political-weapon-covid-19-and-bolsonaro-in-brazil/> (accessed on 29 March 2022). [[CrossRef](#)]
46. Johnson, N.F.; Velásquez, N.; Restrepo, N.J.; Leahy, R.; Gabriel, N.; El Oud, S.; Zheng, M.; Manrique, P.; Wuchty, S.; Lupu, Y. The online competition between pro-and anti-vaccination views. *Nature* **2020**, *582*, 230–233. [[CrossRef](#)] [[PubMed](#)]
47. Whitehead, M.; Taylor, N.; Gough, A.; Chambers, D.; Jessop, M.; Hyde, P. The anti-vax phenomenon. *Vet. Rec.* **2019**, *184*, 744. [[CrossRef](#)] [[PubMed](#)]
48. Kata, A. A postmodern Pandora’s box: Anti-vaccination misinformation on the Internet. *Vaccine* **2010**, *28*, 1709–1716. [[CrossRef](#)] [[PubMed](#)]
49. Helberger, N. The political power of platforms: How current attempts to regulate misinformation amplify opinion power. *Digit. J.* **2020**, *8*, 842–854. [[CrossRef](#)]
50. Sky. Coronavirus: Brazil President Refuses to Ramp Up COVID-19 Lockdown as Facebook Pulls Video. 2020. Available online: <https://news.sky.com/story/coronavirus-brazil-president-refuses-to-ramp-up-covid-19-lockdown-as-facebook-pulls-video-11966279> (accessed on 10 January 2021).
51. Central Intelligence Agency. The World Factbook—Country Comparison—Population. 2018. Available online: <https://www.cia.gov/the-world-factbook/references/guide-to-country-comparisons/> (accessed on 10 January 2021).
52. Oshikawa, R.; Qian, J.; Wang, W.Y. A survey on natural language processing for fake news detection. *arXiv* **2018**, arXiv:1811.00770.
53. Oraby, S.; Reed, L.; Compton, R.; Riloff, E.; Walker, M.; Whittaker, S. And that’s a fact: Distinguishing factual and emotional argumentation in online dialogue. *arXiv* **2017**, arXiv:1709.05295.
54. Aphiwongsophon, S.; Chongstitvatana, P. Detecting fake news with machine learning method. In Proceedings of the 2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Chiang Rai, Thailand, 18–21 July 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 528–531.
55. Singh, V.; Dasgupta, R.; Sonagra, D.; Raman, K.; Ghosh, I. Automated fake news detection using linguistic analysis and machine learning. In Proceedings of the International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation (SBP-BRIMS), Washington, DC, USA, 5–8 July 2017; pp. 1–3.
56. Pérez-Rosas, V.; Kleinberg, B.; Lefevre, A.; Mihalcea, R. Automatic detection of fake news. *arXiv* **2017**, arXiv:1708.07104.

57. Zhang, H.; Fan, Z.; Zheng, J.; Liu, Q. An improving deception detection method in computer-mediated communication. *J. Netw.* **2012**, *7*, 1811. [[CrossRef](#)]
58. Feng, S.; Banerjee, R.; Choi, Y. Syntactic stylometry for deception detection. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Jeju Island, Korea, 8–14 July 2012; pp. 171–175.
59. Zhou, X.; Jain, A.; Phoha, V.V.; Zafarani, R. Fake news early detection: A theory-driven model. *Digit. Threat. Res. Pract.* **2020**, *1*, 1–25. [[CrossRef](#)]
60. Hassan, N.; Arslan, F.; Li, C.; Tremayne, M. Toward automated fact-checking: Detecting check-worthy factual claims by claimbuster. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017; pp. 1803–1812.
61. Reis, J.C.; Correia, A.; Murai, F.; Veloso, A.; Benevenuto, F. Explainable machine learning for fake news detection. In Proceedings of the 10th ACM Conference on Web Science, Boston, MA, USA, 30 June–3 July 2019; pp. 17–26.
62. Kaliyar, R.K.; Goswami, A.; Narang, P.; Sinha, S. FNDNet—a deep convolutional neural network for fake news detection. *Cogn. Syst. Res.* **2020**, *61*, 32–44. [[CrossRef](#)]
63. Kumar, S.; Asthana, R.; Upadhyay, S.; Upreti, N.; Akbar, M. Fake news detection using deep learning models: A novel approach. *Trans. Emerg. Telecommun. Technol.* **2020**, *31*, e3767. [[CrossRef](#)]
64. Yang, Y.; Zheng, L.; Zhang, J.; Cui, Q.; Li, Z.; Yu, P.S. TI-CNN: Convolutional Neural Networks for Fake News Detection. *arXiv* **2018**, arXiv:1806.00749v1.
65. Wang, W.Y. “liar, liar pants on fire”: A new benchmark dataset for fake news detection. *arXiv* **2017**, arXiv:1705.00648.
66. Girgis, S.; Amer, E.; Gadallah, M. Deep learning algorithms for detecting fake news in online text. In Proceedings of the 2018 13th International Conference on Computer Engineering and Systems (ICCES), Cairo, Egypt, 18–19 December 2018; pp. 93–97.
67. Bajaj, S. The Pope Has a New Baby! Fake News Detection Using Deep Learning. CS 224N. 2017, pp. 1–8. Available online: <https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1174/reports/2710385.pdf> (accessed on 24 January 2022).
68. Shu, K.; Cui, L.; Wang, S.; Lee, D.; Liu, H. defend: Explainable fake news detection. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 395–405.
69. O’Brien, N.; Latessa, S.; Evangelopoulos, G.; Boix, X. The Language of Fake News: Opening the Black-Box of Deep Learning Based Detectors. 2018. Available online: <https://cbmm.mit.edu/sites/default/files/publications/fake-news-paper-NIPS.pdf> (accessed on 24 January 2022).
70. Singhania, S.; Fernandez, N.; Rao, S. 3han: A deep neural network for fake news detection. In Proceedings of the International Conference on Neural Information Processing, Guangzhou, China, 14–18 November 2017; Springer: Berlin/Heidelberg, Germany, 2017; pp. 572–581.
71. Thota, A.; Tilak, P.; Ahluwalia, S.; Lohia, N. Fake news detection: A deep learning approach. *SMU Data Sci. Rev.* **2018**, *1*, 10.
72. Huang, Y.F.; Chen, P.H. Fake news detection using an ensemble learning model based on self-adaptive harmony search algorithms. *Expert Syst. Appl.* **2020**, *159*, 113584. [[CrossRef](#)]
73. Roy, A.; Basak, K.; Ekbal, A.; Bhattacharyya, P. A deep ensemble framework for fake news detection and classification. *arXiv* **2018**, arXiv:1811.04670.
74. Vijjali, R.; Potluri, P.; Kumar, S.; Teki, S. Two stage transformer model for COVID-19 fake news detection and fact checking. *arXiv* **2020**, arXiv:2011.13253.
75. Cui, L.; Lee, D. CoAID: COVID-19 Healthcare Misinformation Dataset. *Soc. Inf. Netw.* **2020**, arXiv:2006.00885
76. Wani, A.; Joshi, I.; Khandve, S.; Wagh, V.; Joshi, R. Evaluating Deep Learning Approaches for COVID-19 Fake News Detection. *arXiv* **2021**, arXiv:2101.04012.
77. Glazkova, A.; Glazkov, M.; Trifonov, T. g2tmn at Constraint@ AAI2021: Exploiting CT-BERT and Ensembling Learning for COVID-19 Fake News Detection. *arXiv* **2020**, arXiv:2012.11967.
78. Chen, E.; Lerman, K.; Ferrara, E. Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus twitter data set. *JMIR Public Health Surveill.* **2020**, *6*, e19273. [[CrossRef](#)] [[PubMed](#)]
79. Banda, J.M.; Tekumalla, R.; Wang, G.; Yu, J.; Liu, T.; Ding, Y.; Chowell, G. A large-scale COVID-19 Twitter chatter dataset for open scientific research—An international collaboration. *Epidemiologia* **2021**, *2*, 315–324. [[CrossRef](#)]
80. Kleinberg, B.; van der Vegt, I.; Mozes, M. Measuring emotions in the COVID-19 real world worry dataset. *arXiv* **2020**, arXiv:2004.04225.
81. Li, Y.; Jiang, B.; Shu, K.; Liu, H. MM-COVID: A multilingual and multimodal data repository for combating COVID-19 disinformation. *arXiv* **2020**, arXiv:2011.04088.
82. Galhardi, C.P.; Freire, N.P.; Minayo, M.C.d.S.; Fagundes, M.C.M. Fact or Fake? An analysis of disinformation regarding the COVID-19 pandemic in Brazil. *Ciência Saúde Coletiva* **2020**, *25*, 4201–4210. [[CrossRef](#)] [[PubMed](#)]
83. Pobiruchin, M.; Zowalla, R.; Wiesner, M. Temporal and Location Variations, and Link Categories for the Dissemination of COVID-19-Related Information on Twitter During the SARS-CoV-2 Outbreak in Europe: Inveillance Study. *J. Med. Internet Res.* **2020**, *22*, e19629. [[CrossRef](#)] [[PubMed](#)]
84. Oettershagen, L.; Kriege, N.M.; Morris, C.; Mutzel, P. Classifying Dissemination Processes in Temporal Graphs. *Big Data* **2020**, *8*, 363–378. [[CrossRef](#)]

85. Garcia Filho, C.; Vieira, L.J.E.d.S.; Silva, R.M.d. Buscas na internet sobre medidas de enfrentamento à COVID-19 no Brasil: Descrição de pesquisas realizadas nos primeiros 100 dias de 2020. *Epidemiologia e Serviços de Saúde* **2020**, *29*, e2020191. [[CrossRef](#)] [[PubMed](#)]
86. Ceron, W.; de Lima-Santos, M.F.; Quiles, M.G. Fake news agenda in the era of COVID-19: Identifying trends through fact-checking content. *Online Soc. Netw. Media* **2020**, *21*, 100116. [[CrossRef](#)]
87. Monteiro, R.A.; Santos, R.L.; Pardo, T.A.; De Almeida, T.A.; Ruiz, E.E.; Vale, O.A. Contributions to the study of fake news in portuguese: New corpus and automatic detection results. In Proceedings of the International Conference on Computational Processing of the Portuguese Language, Canela, Brazil, 24–26 September 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 324–334.
88. Silva, R.M.; Santos, R.L.; Almeida, T.A.; Pardo, T.A. Towards automatically filtering fake news in portuguese. *Expert Syst. Appl.* **2020**, *146*, 113199. [[CrossRef](#)]
89. de Souza, M.P.; da Silva, F.R.M.; Freire, P.M.S.; Goldschmidt, R.R. A Linguistic-Based Method that Combines Polarity, Emotion and Grammatical Characteristics to Detect Fake News in Portuguese. In Proceedings of the Brazilian Symposium on Multimedia and the Web, São Luís, Brazil, 30 November–4 December 2020; pp. 217–224.
90. Faustini, P.; Covões, T.F. Fake News Detection Using One-Class Classification. In Proceedings of the 2019 8th Brazilian Conference on Intelligent Systems (BRACIS), Salvador, Brazil, 15–18 October 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 592–597.
91. Cabral, L.; Monteiro, J.M.; da Silva, J.W.F.; Mattos, C.L.C.; Mourao, P.J.C. Fakewhastapp. br: NLP and machine learning techniques for misinformation detection in brazilian portuguese whatsapp messages. In Proceedings of the 23rd International Conference on Enterprise Information Systems, ICEIS, Online, 26–28 April 2021; pp. 26–28.
92. Martins, A.D.F.; Cabral, L.; Mourao, P.J.C.; de Sá, I.C.; Monteiro, J.M.; Machado, J. COVID19. br: A dataset of misinformation about COVID-19 in brazilian portuguese whatsapp messages. In *Anais do III Dataset Showcase Workshop*; SBC: Porto Alegre, Brazil, 2021; pp. 138–147.
93. Martins, A.D.F.; Cabral, L.; Mourao, P.J.C.; Monteiro, J.M.; Machado, J. Detection of misinformation about covid-19 in brazilian portuguese whatsapp messages using deep learning. In *Anais do XXXVI Simpósio Brasileiro de Bancos de Dados*; SBC: Porto Alegre, Brazil, 2021; pp. 85–96.
94. Globo, J.O. Fato ou Fake Coronavirus | Jornal O Globo. 2020. Available online: <https://g1.globo.com/fato-ou-fake/coronavirus/> (accessed on 27 July 2020).
95. Bivar, G.C.C.; de Aguiar, M.E.S.C.; Santos, R.V.C.; Cardoso, P.R.G. COVID-19, the anti-vaccine movement and immunization challenges in Brazil. *Sci. Med.* **2021**, *31*, e39425. [[CrossRef](#)]
96. Rochel de Camargo, K., Jr. Here we go again: The reemergence of anti-vaccine activism on the Internet. *Cadernos de Saúde Pública* **2020**, *36*, e00037620. [[CrossRef](#)] [[PubMed](#)]
97. Jolley, D.; Douglas, K.M. The effects of anti-vaccine conspiracy theories on vaccination intentions. *PLoS ONE* **2014**, *9*, e89177. [[CrossRef](#)] [[PubMed](#)]
98. Nguyen, C.T. Echo Chambers And Epistemic Bubbles. 2020. Available online: <https://www.cambridge.org/core/journals/episteme/article/abs/echo-chambers-and-epistemic-bubbles/5D4AC3A808C538E17C50A7C09EC706F0> (accessed on 22 March 2022).
99. Zhou, L.; Twitchell, D.P.; Qin, T.; Burgoon, J.K.; Nunamaker, J.F. An exploratory study into deception detection in text-based computer-mediated communication. In Proceedings of the 36th Annual Hawaii International Conference on System Sciences, Big Island, HI, USA, 6–9 January 2003; IEEE: Piscataway, NJ, USA, 2003.
100. Zhou, L.; Burgoon, J.K.; Twitchell, D.P.; Qin, T.; Nunamaker, J.F., Jr. A comparison of classification methods for predicting deception in computer-mediated communication. *J. Manag. Inf. Syst.* **2004**, *20*, 139–166. [[CrossRef](#)]
101. Mehrabian, A.; Wiener, M. Non-immediacy between communicator and object of communication in a verbal message: Application to the inference of attitudes. *J. Consult. Psychol.* **1966**, *30*, 420. [[CrossRef](#)]
102. Lynn, T.; Endo, P.T.; Rosati, P.; Silva, I.; Santos, G.L.; Ging, D. A comparison of machine learning approaches for detecting misogynistic speech in urban dictionary. In Proceedings of the 2019 International Conference on Cyber Situational Awareness, Data Analytics And Assessment (Cyber SA), Virtual, 3–4 June 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–8.
103. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer Science & Business Media: Berlin, Germany, 2013.
104. Kowsari, K.; Jafari Meimandi, K.; Heidarysafa, M.; Mendu, S.; Barnes, L.; Brown, D. Text classification algorithms: A survey. *Information* **2019**, *10*, 150. [[CrossRef](#)]
105. Alzamzami, F.; Hoda, M.; El Saddik, A. Light Gradient Boosting Machine for General Sentiment Classification on Short Texts: A Comparative Evaluation. *IEEE Access* **2020**, *8*, 101840–101858. [[CrossRef](#)]
106. Graves, A.; Fernández, S.; Schmidhuber, J. Bidirectional LSTM networks for improved phoneme classification and recognition. In Proceedings of the International Conference on Artificial Neural Networks, Warsaw, Poland, 11–15 September 2005; Springer: Berlin/Heidelberg, Germany, 2005; pp. 799–804.
107. Grandini, M.; Bagli, E.; Visani, G. Metrics for Multi-Class Classification: An Overview. *arXiv* **2020**, arXiv:2008.05756.
108. Williams, N.; Zander, S. *Evaluating Machine Learning Algorithms for Automated Network Application Identification*; Swinburne University of Technology, Centre for Advanced Internet Architectures: Melbourne, VIC, Australia, 2006.
109. Derczynski, L. Complementarity, F-score, and NLP Evaluation. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), Portorož, Slovenia, 23–28 May 2016; pp. 261–266.

110. Méndez, J.R.; Iglesias, E.L.; Fdez-Riverola, F.; Díaz, F.; Corchado, J.M. Tokenising, stemming and stopword removal on anti-spam filtering domain. In Proceedings of the Conference of the Spanish Association for Artificial Intelligence, Málaga, Spain, 22–24 September 2021; Springer: Berlin/Heidelberg, Germany, 2005; pp. 449–458.
111. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
112. Liashchynskyi, P.; Liashchynskyi, P. Grid search, random search, genetic algorithm: A big comparison for nas. *arXiv* **2019**, arXiv:1912.06059.
113. Paixao, M.; Lima, R.; Espinasse, B. Fake News Classification and Topic Modeling in Brazilian Portuguese. In Proceedings of the 2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), Melbourne, Australia, 14–17 December 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 427–432.
114. Kilimci, Z.H.; Akyuz, A.O.; Uysal, M.; Akyokus, S.; Uysal, M.O.; Atak Bulbul, B.; Ekmiş, M.A. An improved demand forecasting model using deep learning approach and proposed decision integration strategy for supply chain. *Complexity* **2019**, *2019*, 9067367. [[CrossRef](#)]