

## Article

# Know an Emotion by the Company It Keeps: Word Embeddings from Reddit/Coronavirus

Alejandro García-Rudolph <sup>1,2,3,\*</sup>, David Sanchez-Pinsach <sup>1,2,3</sup>, Dietmar Frey <sup>4</sup>, Eloy Opisso <sup>1,2,3</sup>, Katryna Cisek <sup>5</sup> and John D. Kelleher <sup>5</sup>

<sup>1</sup> Department of Research and Innovation, Institut Guttmann, Institut Universitari de Neurorehabilitació Adscrit a la UAB, 08027 Badalona, Spain

<sup>2</sup> Universitat Autònoma de Barcelona, Bellaterra (Cerdanyola del Vallès), 08193 Barcelona, Spain

<sup>3</sup> Fundació Institut d'Investigació en Ciències de la Salut Germans Trias i Pujol, 08916 Badalona, Spain

<sup>4</sup> CLAIM Charité Lab for AI in Medicine, Charité Universitätsmedizin Berlin, 10117 Berlin, Germany

<sup>5</sup> Information, Communication and Entertainment Research Institute, Technological University Dublin (TU Dublin), D7 EWV4 Dublin, Ireland

\* Correspondence: alejandropablogarcia@gmail.com

**Abstract:** Social media is a crucial communication tool (e.g., with 430 million monthly active users in online forums such as Reddit), being an objective of Natural Language Processing (NLP) techniques. One of them (word embeddings) is based on the quotation, “You shall know a word by the company it keeps,” highlighting the importance of context in NLP. Meanwhile, “Context is everything in Emotion Research.” Therefore, we aimed to train a model (W2V) for generating word associations (also known as embeddings) using a popular Coronavirus Reddit forum, validate them using public evidence and apply them to the discovery of context for specific emotions previously reported as related to psychological resilience. We used Pushshiftr, quanteda, broom, wordVectors, and superheat R packages. We collected all 374,421 posts submitted by 104,351 users to Reddit/Coronavirus forum between January 2020 and July 2021. W2V identified 64 terms representing the context for seven positive emotions (gratitude, compassion, love, relief, hope, calm, and admiration) and 52 terms for seven negative emotions (anger, loneliness, boredom, fear, anxiety, confusion, sadness) all from valid experienced situations. We clustered them visually, highlighting contextual similarity. Although trained on a “small” dataset, W2V can be used for context discovery to expand on concepts such as psychological resilience.

**Keywords:** COVID-19; social media; Reddit; natural language processing; emotions; resilience

**Citation:** García-Rudolph, A.; Pinsach, D.S.; Frey, D.; Opisso, E.; Cisek, K.; Kelleher, J.D. Know an Emotion by the Company it Keeps: Word Embeddings from Reddit/Coronavirus. *Appl. Sci.* **2023**, *13*, 6713. <https://doi.org/10.3390/app13116713>

Academic Editor: Xiangjie Kong

Received: 20 March 2023

Revised: 20 May 2023

Accepted: 25 May 2023

Published: 31 May 2023



**Copyright:** © 2023 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/>).

## 1. Introduction

Recent statistics show that there are 4.55 billion social media users around the world, equating to 57.6% of the total global population [1].

With the outbreak of the COVID-19 pandemic, social media on platforms such as Reddit [2] has become a critical communication tool for the generation, dissemination, and consumption of information [3].

Therefore, social media analysis is one of the most popular areas of research in recent days [4]. Many studies apply various Natural Language Processing (NLP) techniques to social media content [5]. Out of them, sentiment analysis and topic models are two of the most researched NLP topics, as concluded in a Lancet Digital Health scoping review [3]. Much less studied, word embeddings have been recently reported as a valuable text analysis technology in the pandemic context [6–8].

Understanding the meaning of a word is at the heart of NLP [9]; the approach followed by word embeddings is based on Firth’s notion of “context of situation.” In particular, his famous quotation: “You shall know a word by the company it keeps” [10].

Words that occur in similar contexts are prone to have similar meanings [11]. Firth’s distributional hypothesis is the foundation for the actual word embeddings implementations; one of the most popular is word2vec, developed at Google Labs [12].

Meanwhile, in the field of social sciences, as recently remarked, “Context is everything in Emotion Research” [13]. Few social scientists would refute that context fundamentally shapes psychological experience: our thoughts, feelings, and actions, as well as, to some extent, who we are, and we are all influenced by the context in which we find ourselves.

Context influences cognitions, emotions, and actions in a variety of ways, as well as how these outcomes are seen and understood by others [14,15].

Context is at the core of emotion. “Context is what gives rise to the diversity and depth of human emotional experience and the myriad thoughts and behaviors that stem from such experience” [13].

Existing research indicates that positive emotions support people to cope with stressful situations [16]. This concept is also applicable during times of extended stress, such as the COVID-19 worldwide crisis [17].

Despite the fact that they both share context as a central component, word embeddings have been rarely used in providing context to specific emotions, to the best of our knowledge.

Users of social media platforms, such as Reddit [2], often differ significantly from comparable groups that interact in person. For example, Reddit users are more inclined to discuss issues that they would feel uncomfortable addressing in person [18].

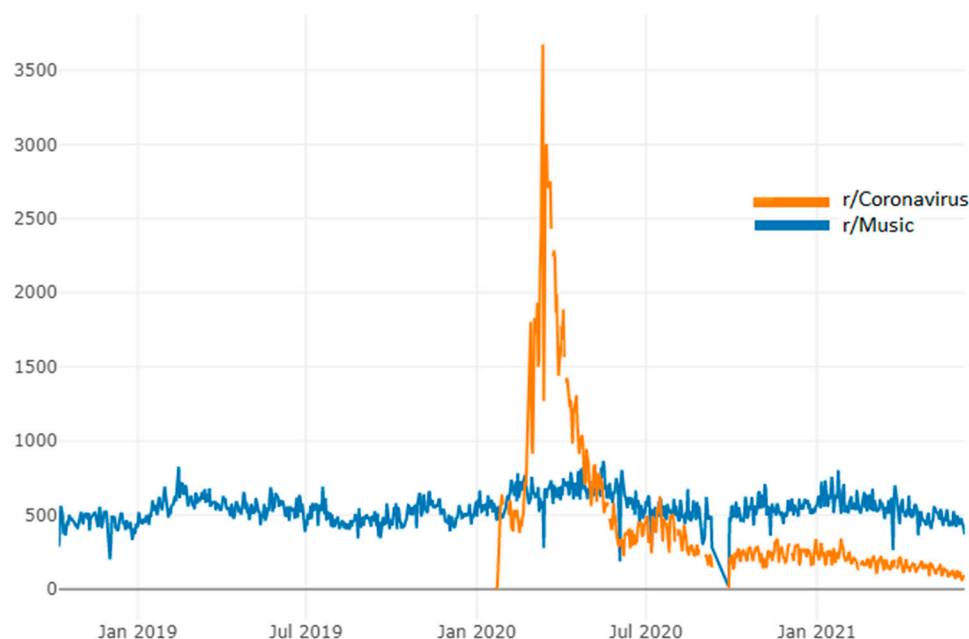
With more than 430 million monthly active users, the primary functionality of Reddit is the exchange of text-based postings through subforums, which are places set aside for users to assemble and communicate with one another on a common interest. The Reddit site name is a play on the words “I read it.” At the end of 2021, there were more than 2.2M Reddit subforums [19] known as subreddits.

Therefore, users can publish posts (also known as submissions) and comments to a number of communities with shared interests called subreddits. Table 1 presents the subreddits related to COVID-19 with the highest number of subscribers. The Rank column shows the absolute position of each subreddit ranked by the number of subscribers as reported by Reddit stats [20]. The r/Music subreddit was included for comparison purposes, as shown in Table 1; r/music was ranked #12 with more than 20M subscribers, one of the most popular subreddits since Reddit was launched [21].

**Table 1.** The top COVID-19 subreddits and their position in the global rank.

Subreddit	Number of Subscribers	Rank	Posts per Day
r/Coronavirus	2,354,224	177	101
r/COVID19	336,253	1357	23
r/CoronavirusUS	140,913	3226	22
r/COVID-19Positive	113,778	3944	29
r/China_Flu	103,456	4261	19
r/CovidVaccinated	27,814	11,271	62
r/Music	20,350,355	12	410

Figure 1 was extracted from subreddit stats [20], and it plots the evolution of r/Coronavirus and r/Music (from January 2019 to July 2021), showing the tremendous increase in posts per day experienced by r/Coronavirus even when compared with one of the most popular subreddits as is the case of r/Music.



**Figure 1.** Number of posts per day for the Music and the Coronavirus subreddits.

The r/Coronavirus subreddit is a curated information platform. As presented in the r/Coronavirus official description [22], “This subreddit seeks to monitor the spread of the disease COVID-19, declared a pandemic by the WHO. This subreddit is for high-quality posts and discussion.” As emphasized in the r/Coronavirus Rules: “There are many places online to discuss conspiracies and speculate, we ask you not to do so here.” Otherwise users get the message: “Your post or comment was removed due to being low quality information” [22]. It is also worth noting that reposts are removed. A repost is a post that is created by taking a post from a while ago and posting it again in the same subreddit. The concept of reposting also covers new posts containing only information that has already been posted [22].

The number of subscribers and posts in the other COVID-19 subreddits are clearly lower and address more specific aspects; therefore, in this work, our data source was r/Coronavirus.

Users submit top-level postings, known as submissions, to each subreddit, while others respond with comments on the submissions. Submissions consist of a title (up to 300 characters) and either a web link or a user-supplied body text; in the latter case, the submission is also known as a self-post, while comments are always made up of a body text.

In this work, we focus on analyzing the titles of Reddit posts. There are two reasons why we believe titles will be a useful basis for NLP analysis.

First, Reddit strongly recommends double-checking the grammar, spelling, and punctuation of the titles: “Read over your submission for mistakes before submitting, especially the title of the submission. Comments and the content of self-posts can be edited after being submitted; however, the title of a post cannot be. Make sure the facts you provide are accurate to avoid any confusion” [23].

Second, Reddit also requests that posters make their titles factual, accurate, and relevant to the content of the post. As remarked in Rediquette: “Please don’t editorialize or sensationalize your submission title, keep your submission titles factual and opinion free. If it is an outrageous topic, share your crazy outrage in the comment section. Do not be vague. Make sure redditors know what they are getting. People do not have time to click on every submission to find out what is inside. Contribute value to the community by writing titles that accurately describe what is being shared. Be relevant. Subreddit

subscribers like to read about specific topics that are related to their subreddit. If your submission is out of place, it will not gain any attention” [23].

Another advantage of focusing on Reddit post titles is that Twitter has increased the available character space from 140 to 280 characters since November 2017, which is very similar to Reddit’s 300 characters limitation of the titles. This provides an opportunity for linguistic comparisons between tweets and Reddit titles.

It is for these reasons that we focused our analysis on the titles of all posts extracted from r/Coronavirus.

A word embedding is a vector-based representation of a word. The vector representing a word can be understood as the coordinates of a word’s position within a multi-dimensional feature space (where the dimensions of the feature space are equal to the size of the vector). Within the vector-based representation, the meaning of a word is encoded by its position within the feature space relative to other words in the space. From a linguistic semantics perspective, the concept of word embedding is related to the distributional hypothesis for Firth [10], which can be paraphrased as “you shall know the meaning of a word by the company it keeps.” The relationship between the distributional hypothesis and word embeddings is that in well-trained word embedding models, words that occur in similar contexts (i.e., that keep the same company) are positioned close to each other in the feature space (i.e., they have similar vector representations).

Word2vec was created, patented, and published in two papers in 2013 by a team of researchers led by Tomas Mikolov at Google to learn word embeddings from a corpus of language [12]. It creates embeddings for the words in a corpus by training a neural network to predict words that co-occur with other words in the corpus.

Word2Vec includes two alternative strategies for training the neural network: Continuous Bag of Words (CBOW) and Skip-gram. In both of them, a preset length window is moved along the corpus. Using the CBOW strategy, at each step, the network is trained to predict the word in the center of the window based on the surrounding words. In the Skip-gram strategy, the network is trained to predict the other words in the window based on the central word. In both strategies, the learning signal for the network (and hence the information that is encoded in the embeddings the network generates) is the likelihood of one word co-occurring in the surrounding context of another word (i.e., within the same window). In the present paper, we use the Skip-gram model, which has shown better performance in semantic tasks [24].

Psychological resilience, as a general term, deals with how people manage stress and how they recover from traumatic events, encouraging constructive growth and promoting an optimistic outlook on the future [25]. Evidence suggests that when resilience-based abilities are applied to people’s lives, they have many advantages (for example, a carry-over effect on other life domains) [26]. Resilience may be improved with deliberate practice; it is not necessary to be born with it [27]. However, within the research community, there is a lack of a unified definition for the concept [28]. This lack of consensus in definition can also be linked to the lack of consensus on how the concept should be operationalized in order to address community disasters [29]. As recently reported [30], positive and negative emotions have varied effects on developing a resilient attitude. People who go through higher levels of positive emotions (i.e., gratitude, compassion, love, relief, hope, calm, or admiration) exhibit a higher degree of resilience, whereas those who feel high levels of negative emotions (i.e., anger, loneliness, boredom, fear, anxiety, confusion, sadness) are associated with poorer resilience.

Typically, large general-purpose corpora (e.g., Wikipedia dumps with 3 billion words [31]) are used to learn word embeddings. Nevertheless, in this work, we hypothesized that word embeddings could be extracted from publicly available social media, using open source software, in sufficient numbers such that the embeddings (1) are relevant to provide meaningful context to specific emotions specifically linked to an ill-defined domain such as psychological resilience (2) verifiable by sound theoretical semantic tests such as the Battig and Montague norm [32] (3) consistent with current re-

lated scientific publications and (4) offering the possibility of providing actionable knowledge to on-field specialists.

Therefore the objectives of this work are to (1) train a model (W2V) for creating word associations (also known as embeddings) using a publicly available dataset (a subreddit on Coronavirus from January 2020 to July 2021, a period where emotions were exacerbated) and open access software (R libraries) able to retrieve meaningful closest terms. (2) Such a W2V model aims to be formally validated using the semantic categorization test by means of an updated and expanded version [33] of the Battig and Montague norm, with 65 categories; for each category, the silhouette coefficient of the model will be computed. As a complementary validation step, the extensive scientific literature is aimed to be included, supporting our findings. (3) We will then run W2V to discover the context for seven specific positive and seven negative emotions recently reported as related to resilience during the COVID-19 pandemic, and (4) such specific context will be supported using related scientific publications.

The article is organized as follows. A literature review is presented in Section 2. Materials and methods are introduced in Section 3. In Section 4, we initially report a descriptive analysis of the sample; we then present the results of our W2V model at three different levels (toy-example analogies, representative terms from a COVID-19 glossary, and resilience related terms) for both the COVID-19 glossary and resilience related terms. We support our findings with extensive scientific literature and then discuss performance using the Battig and Montague evaluation. The discussion and limitations are presented in Section 5. Lastly, in Section 6, we conclude the paper.

## 2. Related Work

For computer scientists and researchers, social media data are valuable assets for understanding people's sentiments regarding current events, especially those related to events with worldwide impacts, such as the COVID-19 pandemic. Therefore, the classification of these sentiments yields remarkable findings. For example, in one of the earliest related publications, Rajput and colleagues [34] classified (negative, positive, and neutral) tweets based on word-level, bi-gram, and tri-gram frequencies to represent word rates by power law distribution and applied the Python built-in package TextBlob to perform sentiment analysis. Samuel and colleagues [35] proposed machine learning models (naïve Bayes and logistic regression) to categorize sentiment tweets into two classes (positive and negative). Similarly, Aljameel et al. [36] analyzed a large Arabic COVID-19-related tweets dataset, applying uni-gram and bi-gram TF-IDF with SVM, naïve Bayes, and KNN classifiers to enhance accuracy. Muthausami et al. [37] classified the tweets into three classes (positive, neutral, and negative). They utilized different classifiers, such as random forest, SVM, decision tree, naïve Bayes, LogitBoost, and MaxEntropy. More recently, Jalil and colleagues [38] classified positive, negative, and neutral tweets using various feature sets and XGBoost (eXtreme Gradient Boosting) classifier. The authors of Rustam et al. [39] proposed a COVID-19 tweets classification approach based on a decision tree, XGBoost, extra tree classifier (ETC), random forest, and LSTM. Similarly, Dangi et al. [40] proposed a novel approach known as Sentimental Analysis of Twitter social media Data (SATD) based on five different machine learning models (logistic regression, random forest classifier, multinomial NB classifier, support vector machine, and decision tree classifier)

Rahman et al. [41] explored the performance of ensemble machine learning classifiers for sentiment analysis of COVID-19 tweets from the United Kingdom. Es-Sabery et al. [42] applied MapReduce opinion mining for COVID-19-related tweets classification using an enhanced ID3 decision tree classifier.

Basiri et al. [43] presented a model that combines five models such as naïve Bayes support vector machines (NBSVM), FastText, DistilBERT, CNN, and bidirectional gated recurrent unit (BiGRU) on COVID-19 tweets in eight highly affected countries. Ibrahim et al. [44] proposed a hierarchical Twitter sentiment model (HTSM) to show people's opin-

ions in short texts. Bonifazi et al. [45] proposed a novel approach for investigating the COVID-19 discussions on Twitter through a multilayer network-based model. It yielded the identification of influential users, which is much more important to analyze and can provide more valuable information.

Naseem et al. [46] correspondingly proposed the use of various pre-trained embedding representations—FastText, GloVe, Word2Vec, and BERT—to extract features from a Twitter dataset. Furthermore, for the classification, they applied deep learning methods Bi-LSTM and several classical machine learning classifiers, such as SVM and naïve Bayes.

Yan et al. [47] reported public sentiment toward COVID-19 vaccines across Canadian cities by analyzing comments on Reddit. In order to identify significant latent topics and classify sentiments in COVID-19-related English comments between January and March 2020, Jelodar et al. examined 563,079 comments from Reddit [48]. Lai et al. [49] analyzed 522 comments from a Reddit Ask Me Anything session about COVID-19. Reddit posts evaluated in this study were manually coded by two authors of this paper.

Pal et al. [50] showed that new knowledge could be captured and tracked using the temporal change in word embeddings from the abstracts of COVID-19 published articles. They found that thromboembolic complications were detected as an emerging theme as of August 2020. A shift toward the symptoms of long COVID complications was observed in March 2021, and neurological complications gained significance in June 2021.

Jha et al. [51] observed that the word2vec model performed better than the GloVe model on a COVID-19 Kaggle dataset. Another point highlighted by this work is that latent information about potential future discoveries was significantly contained in past papers and publications.

Batzdorfer et al. [52] used word embeddings to distinguish non-conspiracy theory content from conspiracy theory-related content and analyzed which element of conspiracy theory content emerged during the pandemic.

Didi et al. [6] proposed a tweets classification approach (negative, positive, and neutral) based on a hybrid word embedding method, combining several widely used techniques, such as TF-IDF, word2vec, Glove, and FastText, to represent posts.

Bhandari et al. [53] proposed a deep learning model with stacked word embeddings to the multi-class classification problem for three and five classes (extremely negative, negative, neutral, extremely positive, and positive). It outperformed the individual static pre-trained embedding representation, classical machine, and deep learning approaches.

To our knowledge, no previous analysis applied word embeddings to extract knowledge from Reddit to provide context about specific emotions involved in psychological resilience during the pandemic. Acute crisis and loss events, disruptions in many facets of life, continuous multi-stress problems, and always-changing conditions made the COVID-19 pandemic a perfect storm of stressors. The rapid spread of COVID-19 during the 2020-2021 period, when emotions were exacerbated [54], created a unique opportunity to extract knowledge about resilience in the face of global adversity, yet to be explored using NLP. We believe that a better understanding of resilience is important in developing strategies to cultivate and promote resilience.

### 3. Methods

Our research included different sequential phases starting with data collection from publicly available Reddit titles from the R/Coronavirus subreddit, data cleaning using open access R libraries, an initial descriptive analysis of the available data, word2vec model training, the formal model validation using semantic categorization test and visualization using hierarchical clustering and heatmaps. Each of them is described in this section.

### 3.1. Data Collection

Data from Reddit were obtained via pushshift.io through the pushshift.io API (Pushshift, 2023) [55]. In order to collect and distribute Reddit datasets for research purposes, academics can use Pushshift.io, a website that keeps all publicly accessible Reddit submissions and comments. Pushshift.io has been used in a large number of publications in related research (e.g., Lama et al. [56]). In this work, the pushshiftR R package [57] was used as a wrapper for the pushshift.io API.

### 3.2. Data Cleaning

The quantda R library [58] was used to create the final sample for analysis. The data cleaning process included lemmatization (where the phrases “dog,” “dogs,” and “dog’s” are all changed to “dog”), nonprintable character removal (such as emojis), and basic normalizing (such as removing punctuation and lowercasing all text).

All analyses used are publicly available, anonymized data and comply with Reddit’s terms of service, usage rules, and privacy guidelines. They were also carried out with institutional review board clearance from the authors’ institutions.

### 3.3. Descriptive Initial Analysis

For descriptive analysis, we first processed the data into the tidy text format as one token (word) per row. The process of breaking text into tokens is known as tokenization. This one-token-per-row structure differs from how text is commonly kept in current studies (e.g., in a document-term matrix). For tidy text pre-processing, we used the tidytext [59], dplyr [60], ggplot2 [61], and broom [62] R packages.

In order to determine if the frequency of each word is rising or decreasing over time, we fitted a model (logistic regression) using the broom R package. Then, each term has a growth rate (represented by an exponential term) associated with it.

In the Supplementary material Figure S1, we present the number of titles per week. We confirm that the distribution is quite similar to the plot provided by the official Reddit statistics presented in Figure 1.

Figure S2 shows the most frequent words (after excluding COVID-19, Coronavirus, and pandemic, which due to their highest frequency, make all other terms not visible if put together in the same plot with all other terms). The top 10 are people, vaccine, China, positive, health, home, masks, world, death, and Trump.

Figure S3 shows the terms with the steepest increase in frequency. The highest one is for Donald Trump, right before the day of the Presidential Election in the United States (3 November 2020), with the highest decrease after it. When visualizing all four sub-plots in Figure S2, shown from left to right and from top to bottom, it can be seen that each of them refers to a specific aspect of this pandemic, each of them with special relevance at different time points: lockdown at the early stage, masks and Trump at intermediate stages, and vaccine increasing steadily until the final stages.

In Figure S4, we present a word cloud created using all the titles containing the term “stress.”

### 3.4. Model Training: Word2vec

We applied the wordVectors [63] R package to train the word2vec model. It runs the original C code for word2vec [12].

A metric of the degree of similarity between two embedding vectors for the two words is provided to measure how similar the two words are. Given two vectors  $u$  and  $v$ , cosine similarity is defined as follows [12]:

$$\text{CosineSimilarity}(u, v) = \frac{u \cdot v}{\|u\|_2 \|v\|_2} = \cos \theta \quad (1)$$

where  $u.v$  is the dot product (or inner product) of two vectors,  $\|u\|_2$  is the norm (or length) of the vector  $u$ , and  $\theta$  is the angle between  $u$  and  $v$ .

The cosine distance is defined as the inverse of the cosine similarity; the shorter the cosine distance, the more similar the two vectors (words).

### 3.5. Model Validation: Semantic Categorization Test

We measured the capacity of the W2V model to represent the semantic categories based on the Battig and Montague category norms, which have been applied by researchers in several fields in over 1600 publications in more than 200 different journals [33]. In this work, we use Van Overschelde's [33] expanded and updated version of the Battig and Montague original norms.

In order to measure how well a word  $i$  is grouped in relation to the other words in its semantic category, we used the Silhouette Coefficients,  $s(i)$ , defined as:

$$s(i) = (b(i) - a(i)) / \max\{a(i), b(i)\}$$

where  $a(i)$  is the mean distance of word  $i$  with all other words within the same category, and  $b(i)$  is the minimum mean distance of word  $i$  to any words within another category (i.e., the mean distance to the neighboring category). Therefore, silhouette coefficients measure how close a word is to other words within the same category compared to words of the closest category [64].

### 3.6. Model Visualization: Hierarchical Clustering and Heatmaps

We used the superheat R package [65] to visualize the word vectors (obtained from Word2vec), highlighting contextual similarity. "The rows and columns are ordered based on a hierarchical clustering and are accompanied by dendrograms describing this hierarchical cluster structure" [65].

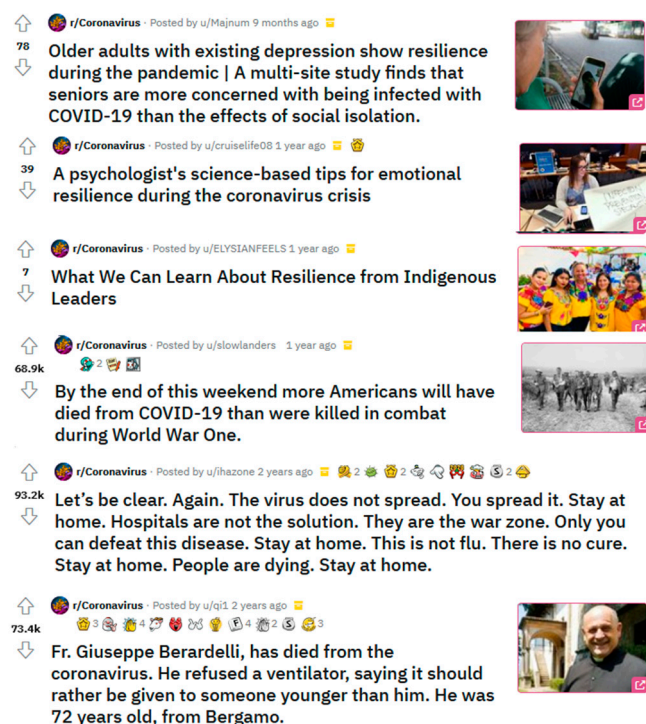
## 4. Results

### 4.1. Sample Description

We collected all 374,421 titles submitted by 104,351 different Redditors to the r/Coronavirus subreddit between 20 January 2020 and 14 July 2021.

In Figure 2, we show representative examples of the collected titles, the top 3 containing the term "resilience" and the bottom three randomly selected.





**Figure 2.** Examples of the collected titles containing the term “resilience” (top 3) and randomly selected (bottom 3).

#### 4.2. A 3-Steps Validation of the Word2vec Embeddings

The `train_word2vec` function of the `wordVectors` R package was used to obtain the model (W2V) once the data had been generated. The following settings were used: “vectors = 200, threads = 4, window = 12, iter = 5, negative\_samples = 0”. These parameters have been applied by the *wordVectors* authors in related research [63].

We performed a three-step validation of W2V as in previous related research [66]. We utilized a subset of the original Mikolov article analogies [12] for the first one.

In NLP, the task of finding a word analogy is represented as “a is to b as c is to \_\_\_\_.”

The classic Mikolov example is: king is to man as woman is to \_\_\_\_—also represented as king – man + woman = ?

The human brain can recognize that the answer is the word ‘queen’. However, for a machine to understand this pattern and fill in the blank with the most appropriate word requires a lot of training using a huge corpus (for example, the whole of Wikipedia; in our case, we are using only the obtained 374,421 titles from r/Coronavirus).

Using our obtained model (namely W2V), the example analogy is represented as:  $W2V(\text{“king”}) - W2V(\text{“man”}) + W2V(\text{“woman”}) = ?$

We obtained promising results (as presented in Table 2) for several analogies from previous research [66], for example:

Analogy: brother – sister + husband = ?

Answer: wife (0.5985)

The number in brackets is the cosine distance between the vector embedding for the term ‘wife’ and the vector that is the result of the operations on the left-hand side of the equation.

**Table 2.** A subset of analogies from previous research [66] and the obtained results.

Category	Closest Terms (Cosine Distance)
paris – france + italy = ?	rome (0.584), milan (0.510)
brother – sister + husband = ?	wife (0.598)
dad – mom + father = ?	mother (0.546), family (0.569)
she – he + girl = ?	boy (0.375)
his – her + boy = ?	girl (0.570), schoolgirl (0.604)
she – he + mother = ?	father (0.373), husband (0.403)
boy – girl + man = ?	woman (0.553)
doctor – hospital + teacher = ?	school (0.577), teen (0.548)
cnr – news + netflix = ?	film (0.640), movies (0.692)
iphone – apple + android = ?	ios(0.406), tablet (0.4760), app (0.487)
moscow – putin + nyc	Blasio * (0.619), brooklyn (0.581)
young – teen + old	64 (0.633), aged (0.563)

\* Bill de Blasio is an American politician serving as the 109th Mayor of New York City since 2014.

As the second step of W2V validation, from a representative list of specific terms related to COVID-19, we run our W2V model on each of them (for example, the term “anosmia”) to identify its three closest terms using the following command:

```
nearest_to(W2V[["anosmia"]],3) = ?
```

As a result, we obtained the following set of the three closest terms to “anosmia”:

```
{olfactory (0.463); parkinson (0.459); aspirin (0.496)}
```

In Table 3, we present the closest terms retrieved by our model and their cosine distances to several COVID-19 representative terms of a known COVID-19 glossary [67]. We proceeded through the closest terms and identified related publications and evidence supporting them, noting the high relevance of all the discovered terms in order to demonstrate the capacity of our W2V model to uncover relevant related terms (Table 3).

**Table 3.** Definitions extracted from COVID-19 Canadian Glossary and our obtained closest terms.

Glossary Term	Glossary Definition	Closest Terms
ARDS	Acute respiratory distress syndrome	remestemcel (0.364) [68], glucose (0.461) [69], epithelium (0.461) [70], anticoagulant (0.481) [71]
anosmia	The complete or partial loss of the sense of smell.	olfactory (0.463) [72] parkinson (0.459) [73], aspirin (0.496) [74],
antibody	A protein that is produced in response to the introduction of an antigen in an organism	monoclonal (0.436) [75], regeneron (0.478) [76], serological (0.475) [77], bamlanivimab (0.539) [78]
antiviral	Medication used for treating viral infections	favipiravir (0.341) [79], remdesivir (0.344) [80], heparin (0.379) [81], interferón (0.385) [82], ritonavir (0.435) [83]

In Table 3, we show the closest terms retrieved by our model and their cosine distances to representative definitions from the initial terms of the glossary (terms starting with the ‘A’ letter). For each of the terms identified by our trained model, we included relevant published scientific literature. For example, the first term in the glossary was

“ards” (acute respiratory distress syndrome); our model retrieved Remestemcel, and its cosine distance was 0.364. We referenced Mahendiratta et al. [68] because in their systematic review of Stem cell therapy in COVID-19, results on Remestemcel were recently reported. Similarly, for glucose, we referenced Lazzeri et al. [69] work, where they address the prognostic role of hyperglycemia and glucose variability in COVID-related acute respiratory distress, similarly, for all other terms in Table 3.

As the third step of W2V validation, we identified the closest terms to “resilience.” Then we searched for all appearances of “resilience” in all 374,421 titles and identified the titles with the highest upvotes. We present them in Figure 2.

In Figure 2 (top 3 titles), we present the most upvoted titles, which explicitly include the term “resilience.” Therefore, we used W2V to search for the closest terms to: resilience appearing in the same context with “older” with “indigenous” and with “tips.” The obtained closest terms are presented in Table 4. We went through all the closest obtained terms and identified related publications and evidence remarking on the high relevance of all the identified terms.

**Table 4.** Resilience-related terms and our obtained closest terms.

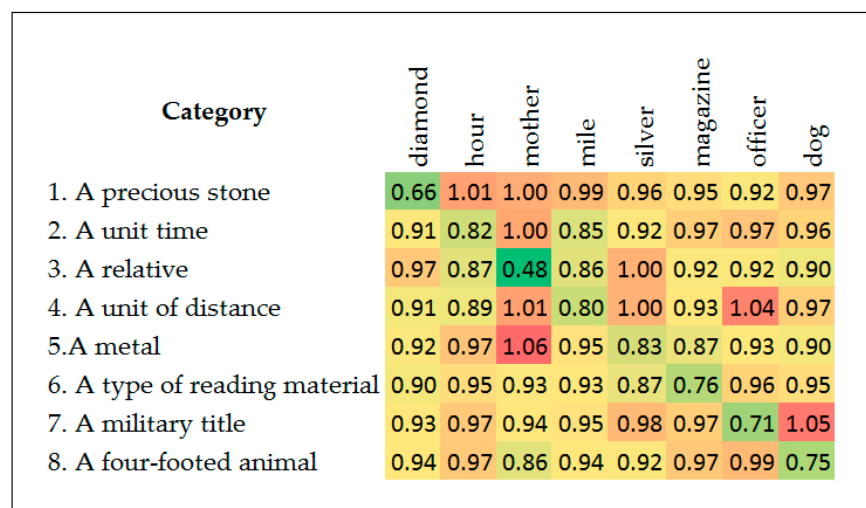
Search Term	Closest Terms
resilience	wellbeing (0.569) [84], pessimism (0.611) [85], psychological (0.586) [85],
resilience + tips	mindfulness (0.580) [86], telehealth (0.588) [87], bedtime (0.577) [88], hobbies (0.546) [89]
resilience + older	addiction (0.570) [90], stress (0.588) [91], disability (0.588) [92], resentment (0.598) [93], depressive (0.617) [94]
resilience + indigenous	communities (0.520), tribe (0.565), minority (0.618), dignity (0.618), unequal (0.622), unicef (0.632), disparities (0.624) [95]

For example, as shown in Table 4, for “resilience” and “older,” we identified several closest terms and included in Table 4 different publications addressing such aspects, e.g., addiction [90], stress [91], disability [92], resentment [93], and depression [94].

#### 4.3. Semantic Categorization Test

For each of the first 65 semantic categories of the updated version of the Battig and Montague norm [33], we calculated the silhouette coefficients. The complete list of all the terms included in each category as well as distances and silhouette calculations, is presented in Supplementary material Table S1. A representative screenshot of the distances from the first eight semantic categories to representative terms is presented in Figure 3. For example, the first semantic category is “1. A precious stone”, as detailed in Table S1. It is integrated into four terms (diamond, ruby, gold, and gem). We run our W2V model to calculate the distances from a representative term from each category to all the other terms. Therefore, as shown in Figure 3, the mean distance from the “diamond” term to all other terms in the “1. A precious stone” category is 0.66. Meanwhile, it is 1.01 to the “2. A unit time” category represented by the “hour” term, it is 1.00 to the “3. A relative” category represented by the “mother” term, and so forth. Therefore, Figure 3 represents such

distances as a heatmap, with greener values to the closest distances. It can be seen that for each term, for every semantic category, the closest distances are to those terms related to the category where the term belongs, therefore showing encouraging results.



**Figure 3.** Heatmap representation of the mean distances between the first 8 semantic categories and their representative terms.

Table 5 presents the highest silhouette values calculated in Supplementary material Table S1. When analyzing the lower Silhouette scores, we identified remarkable reasons for the miscategorization of the terms. For example, as presented in Figure 3, the mean distance from the “diamond” term to all other terms in the “1. A precious stone” category is 0.66, but as shown in Table S1, when considering the “51. A type of ship/boat” category, represented by the “cruise” term, such mean distance is 0.55, remarkably lower. A possible explanation for this is the existence of the Diamond Princess Cruise, which is mentioned in some of the Reddit titles used for training our W2V model.

**Table 5.** Top silhouette values obtained for 10 semantic categories of the updated version of the Battig and Montague norm.

Category	s
29. A sport	0.495
3. A relative	0.329
54. A city	0.233
55. A state	0.231
10. A color	0.169
58. A type of car	0.163
49. A disease	0.154
27. An occupation or profession	0.142
7. A military title	0.139
40. A science	0.137

#### 4.4. Context for Positive and Negative Emotions

In Table 6, we present a list of specific positive emotions (gratitude, compassion, love, relief, hope, calm, and admiration) [30]. We ran our W2V model for each of them and identified several closest terms, providing the context where such emotions took place.

**Table 6.** Positive emotions and their obtained closest terms.

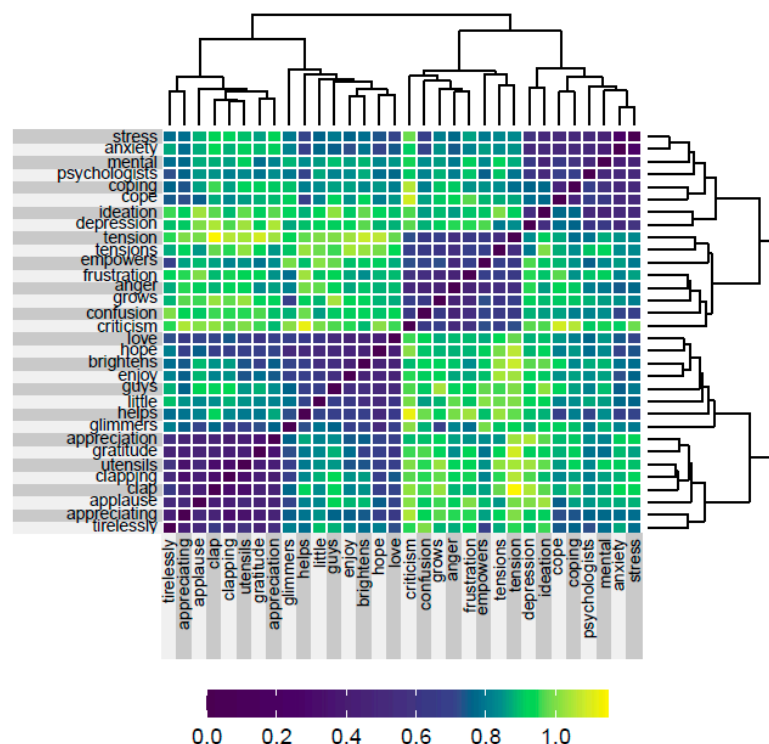
Search Term	Closest Terms
gratitude	paramedical (0.545), doctors (0.495), appreciation (0.368), selflessly (0.498), tirelessly (0.453), heroes (0.503), honor (0.516), tribute (0.540), hardworking (0.555), flashmobs (0.564)
compassion	dalai (0.684), lama (0.685), empathy (0.657), empathetic (0.662), mindfulness (0.633)
love	share (0.400), enjoy (0.440), friends (0.489), wish (0.519), god (0.526), smile (0.528), constructive (0.555), entertain(0.525)
relief	funds (0.325), aid (0.339), package (0.309), fund (0.341), trillion (0.394), billion (0.414), loan (0.409), liquidity (0.414), payments (0.429), payers (0.397), tax (0.436)
hope	Love(0.423), enjoy (0.477), brightens (0.471), help(0.513), inspire(0.517), smile (0.522), laugh (0.536), humor (0.569), fun (0.573), funny (0.573)
calm	Listen(0.532), sleep(0.435), meditation(0.544), Roads(0.521), streets(0.546), eerie(0.656), emptiness(0.668), scary (0.575),,, panic (0.567), nerves (0.546), keep(0.654),
admiration	clapping (0.431), clap (0.455), applause (0.455), balconies (0.484), applauding(0.522), windows (0.472), cheering (0.456), frontline (0.466), healthcare (0.532),

Similarly, Table 7 presents the list of negative emotions [30] (anger, loneliness, boredom, fear, anxiety, confusion, sadness) and their closest terms retrieved using W2V.

**Table 7.** Negative emotions and the obtained closest terms.

Search Term	Closest Terms
anger	frustration (0.471), confusion (0.474), tension (0.555), chaos (0.592), dishonesty (0.630), hostility (0.617), bureaucracy (0.625), fear (0.608), drought (0.579), outcry (0.598), outrage (0.618),
loneliness	profound (0.607), addiction (0.635), neuropsychiatric (0.630), opioid (0.658)
boredom	spotify (0.615), playlists (0.599), song (0.593), halo (0.596), fortnite (0.626), meditation (0.615), illustration (0.631), piano (0.632),
fear	conspiracies (0.587), xenophobia (0.612), racism (0.621), burnout (0.623), starving (0.636), sadness(0.532)
anxiety	stress(0.251), depression(0.431), meditation (0.578), obsessive (0.532), ideation (0.511), cope(0.529), coping (0.514), tips (0.570)
confusion	anger (0.474), frustration (0.546), chaos (0.543), distrust (0.561), tension (0.532), worries (0.522), doubts (0.533)
sadness	Disbelief (0.433), downfall (0.544), dislike (0.541), downvotes(0.541), fear(0.532), boredom (0.533), together(0.544), spinning(0.541)

Figure 4 graphically shows a dendrogram for the closest terms to two positive emotions (hope and gratitude) and two negatives (anger and anxiety) presented as clusters of the most similar closest terms. The darkest the color in the heatmap, the closest are the two terms; therefore, three clear clusters emerge in the heatmap diagonal.



**Figure 4.** Dendrograms and heatmap for the closest terms to two positive (hope and gratitude) and two negative (anger and anxiety) emotions.

## 5. Discussion

In this study, we proposed social media (particularly a Reddit subforum) as a connection between word associations (also known as embeddings) and emotion research. Although they both share context as a critical component, to our best knowledge, word embeddings have rarely been used in the field of emotion research. Furthermore, COVID-19 created a unique opportunity for doing it.

Therefore, we trained a model for producing word embeddings using a publicly accessible dataset (a Coronavirus subreddit) and open-source tools (R libraries) capable of retrieving relevant content (closest words). This content was formally validated using a standard tool and supported by public evidence (scientific publications), and applied to the discovery of context for seven specific positive and seven negative emotions recently reported as related to resilience during the COVID-19 pandemic.

Our results confirmed our three initial hypotheses: word embeddings may be recovered in sufficient numbers from public domain-specific social media for the embedding to (1) be relevant to offer meaningful context to specific emotions, (2) be verifiable by sound theoretical semantic tests such as the Battig and Montague norm, and (3) be consistent with recent related publications, in spite of working with a relative “small” number of Reddit titles.

In relation to our fourth hypothesis (provide actionable knowledge to on-field specialists), current research reporting on the COVID-19 pandemic concluded that developing a resilient mentality differs depending on whether positive or negative emotions are present. Higher levels of positive emotions are correlated with higher levels of resilience, whereas high levels of negative emotions are associated with lower levels of resilience.

ience [30]. We associated seven positive and seven negative emotions to experienced situations. Specialists could therefore promote actions encouraging participation in activities related to positive emotions. For example, as shown in Table 6, “gratitude” and “admiration” were shown by means of activities taking place worldwide. People congregated on balconies while confined to their apartments to acclaim medical personnel working on the front lines, as well as to sing or take part in impromptu flash mobs [96]. Calm and compassion were associated with meditation and mindfulness. Hope was associated with humor, smiling, laughing, fun, and funny.

When analyzing negative emotions, we found racism and xenophobia mainly related to fear. Globally, migrants and minority groups were disproportionately affected by racism and xenophobia linked to COVID-19 [97]. They have an especially negative effect on people who already experience overlapping social, economic, and health-related vulnerabilities. They intensify current patterns of discrimination and unfairness. Minority groups in both the United States and Europe have endured discrimination and hate crimes. [98,99]. Anger was mainly related to frustration, bureaucracy, and confusion as in related research (e.g., Selman et al. [100]); loneliness was associated with addictions, while boredom was related to specific activities to overcome it, such as meditation, illustration, piano, Spotify, playlists or videogames (Halo, Fortnite).

Several recent studies addressed social media (particularly Reddit) during the pandemic. For example, Gozzi et al. analyzed collective responses to media coverage. They performed mixed-methods analysis on web-based news articles, YouTube videos, English user posts and comments on Reddit, and views of Wikipedia pages related to COVID-19. They concluded that “collective attention was mainly driven by media coverage rather than epidemic progression [101]”. Compared to other social media platforms, Reddit users were generally more concerned about health, data related to the new disease, and interventions needed to stop its spread [101]. In order to identify significant latent topics and classify sentiments in COVID-19-related English comments between January and March 2020, Jelodar et al. examined 563,079 comments from Reddit [48]. Lai et al. [49] analyzed 522 comments from a Reddit Ask Me Anything session about COVID-19 on 11 March 2020. Most posts addressed symptoms, followed by prevention recommendations. COVID-19 symptoms were also the most requested topic suggested by users for further discussion.

Word2vec has been scarcely used in small corpora. García-Rudolph et al. [66] analyzed 96,314 Reddit comments posted in r/disability from February 2009 to December 2019 by 10,411 Redditors. The highest reported silhouette value after the semantic categorization test was  $s = 0.562$  for the “3.A relative” category. Meanwhile, in our case, our highest silhouette value was  $s = 0.495$  for the “29.A sport” category. In the “29.A sport” category, their reported silhouette was  $s = 0.475$ . Their top six higher silhouette values were reported for the following categories: 3. A relative, 29. A sport, 43. A vegetable, 10. A color, 55. A state and 49. A disease. In our case, the top six silhouette values were reported for 29. A sport, 3. A relative, 54. A city, 55. A state, 10. A color, and 58. A type of car. Therefore, very similar semantic categories yielded the highest silhouette scores for both studies. Nevertheless, in our case, we collected 374,421 titles (not comments) submitted by 104,351 users (ten times more users) to the Reddit/Coronavirus forum during a ten-times shorter period.

In another study applying word2vec in small corpora using the semantic categorization test, Stetten, the study included 37 k and 140 k documents to analyze and disambiguate the content of dreams [102]. This research area addresses questions such as “How do gender, cultural background, and waking life experiences shape the content of dreams?”. To our knowledge, no previous work studied Reddit submission titles considering word embeddings in order to expand on the concept of resilience. We offer a tool for identifying terms of interest that can be addressed to practitioners in the field of psychology and social work.

A number of limitations to this study need to be highlighted. The analyzed sample was not meant to be exhaustive or representative of all titles posted by everyone living in any specific region during the period under study. It included all titles from only one of the COVID-19 subreddits; therefore, we did not include data from other subreddits addressing specific COVID-19 aspects (e.g., CovidVaccinated or COVID-19Positive). Nevertheless, r/Coronavirus was by far the subreddit with a higher number of subscribers and posts. It has been the most active subreddit during the period under study (between 20 January 2020 and 14 July 2021). We did not include comments in our analysis. We included only submissions' titles. The length limit in Reddit comments is 40,000 characters, more than 100 times larger than the titles' limit (300 characters). Therefore including comments would involve a different analysis, with different hypotheses, which is left as future work.

The potential impact of the data-cleaning process needs to be mentioned as another limitation, particularly in terms of the context of the text. For example, by removing emojis and other non-printable characters, we might have been removing some contextual information that could be relevant to understanding the sentiments or emotions. For example, Li et al. [103] presented an approach to classify microblog review sentiments that included emojis with an emoji-text-incorporating bi-LSTM (ET-BiLSTM) model. Their results showed that ET-BiLSTM enhances the performance of sentiment classification.

Another aspect of Reddit worth to be analyzed, not included in this study, involves NSFW (Not Safe For Work) posts. This term refers to user-submitted content not suitable to be viewed in public or in professional contexts. The phenomenon of NSFW posts on Reddit has been very little investigated, although it is very common in this social medium [104].

Other relevant factors to mention as limitations to our study include geographic location, spatial trajectory, or the time of day a submission was posted. Such factors, as noted by Padilla et al. [105] and Gore et al. [106], are relevant in social media. Geographic aspects were not analyzed in our study, but Reddit is most popular in the U.S., with American users far outnumbering those from any other country at 54% of Reddit users. After the U.S., the United Kingdom has the second-highest share of data traffic with 8%, while Canada ranks third with 6.4%. Reddit is most popular with young adults aged 25 to 34, who comprise more than half of the site's users. Nevertheless, there are also a large number of middle-aged users on Reddit. Previous studies have found that 33% of users are between the ages of 30 and 49, suggesting that Reddit is a viable platform for reaching both young and middle-aged adults. More than two-thirds of Reddit users are men who are particularly active on the site [107]. Compared to people living in rural areas, urban and suburban residents use Reddit much more frequently. Gozzi et al. also pointed out that Reddit has developed into a self-referential community, reinforcing the site's propensity to concentrate on its own content rather than outside sources [101].

## 6. Conclusions

This study opens up interesting opportunities for exploration and discovery using, for the first time, a word2vec model trained with a small Coronavirus dataset of Reddit titles leading to immediate and accurate terms that can be used to expand our knowledge on specific concepts such as resilience, by identifying the context in which they take place. We presented a step forward in developing a tool that can be used by practitioners in the field of psychology or social work for identifying terms of interest describing the context in which specific positive and/or negative emotions related to psychological resilience took place. These may support clinicians in specific situations where individuals can be encouraged to get involved or promote positive emotions related to psychological resilience.



**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/app13116713/s1>, Figure S1: Number of titles per week of the r/Coronavirus subreddit; Figure S2: Top 50 most frequent words; Figure S3: Terms with the steepest increase in frequency; Figure S4: Wordcloud of all titles containing the “stress” term; Table S1: Semantic categorization test.

**Author Contributions:** Methodology: A.G.-R., D.S.-P., J.D.K. and K.C.; Software: A.G.-R., D.S.-P. and K.C.; Validation: E.O. and K.C.; Formal analysis: J.D.K. and A.G.-R.; Investigation: A.G.-R., D.S.-P. and E.O.; Resources: D.F. and E.O.; Data curation: A.G.-R. and D.S.-P.; Writing—original draft: A.G.-R. and D.S.-P.; Writing—review and editing: J.D.K. and E.O.; Visualization: A.G.-R. and D.S.-P.; Supervision: E.O., J.D.K. and D.F. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was partially funded by PRECISE4Q Personalized Medicine by Predictive Modelling in Stroke for Better Quality of Life—European Union’s Horizon 2020 research and innovation program under grant agreement No. 777107.

**Institutional Review Board Statement:** All analyses relied on public, anonymized data; adhered to the terms and conditions, terms of use, and privacy policies of Reddit; and were performed under Institutional Review Board approval from the authors’ institution.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

**Acknowledgments:** Special thanks to Olga Araujo from Institut Guttmann—Documentation department for her continuous support of our bibliography requests.

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

## Abbreviations

API	Application Program Interface.
ARDS	Acute Respiratory Distress Syndrome
CVST	Cerebral Venous Sinus System
CBOW	Continuous Bag of Words
EMA	European Medicines Agency
IgM	Immunoglobulin M
MHRA	Medicines Healthcare products Regulatory Agency
NLP	Natural Language Processing

## References

1. Melton, C.A.; White, B.M.; Davis, R.L.; Bednarczyk, R.A.; Shaban-Nejad, A. Fine-tuned Sentiment Analysis of COVID-19 Vaccine-Related Social Media Data: Comparative Study. *J. Med. Internet Res.* **2022**, *24*, e40408. <https://doi.org/10.2196/40408>.
2. Reddit—Dive into Anything. Founded: June 23, 2005, Medford, Massachusetts, United States. Available online: <https://www.reddit.com/> (accessed on 19 March 2023).
3. Tsao, S.F.; Chen, H.; Tisseverasinghe, T.; Yang, Y.; Li, L.; Butt, Z.A. What social media told us in the time of COVID-19: A scoping review. *Lancet Digit. Health* **2021**, *3*, e175–e194. [https://doi.org/10.1016/S2589-7500\(20\)30315-0](https://doi.org/10.1016/S2589-7500(20)30315-0).
4. White, B.M.; Melton, C.; Zareie, P.; Davis, R.L.; Bednarczyk, R.A.; Shaban-Nejad, A. Exploring celebrity influence on public attitude towards the COVID-19 pandemic: Social media shared sentiment analysis. *BMJ Health Care Inform.* **2023**, *30*, e100665. <https://doi.org/10.1136/bmjhci-2022-100665>.
5. Al-Garadi, M.A.; Yang, Y.C.; Sarker, A. The Role of Natural Language Processing during the COVID-19 Pandemic: Health Applications, Opportunities, and Challenges. *Healthcare* **2022**, *10*, 2270. <https://doi.org/10.3390/healthcare10112270>.
6. Didi, Y.; Walha, A.; Wali, A. COVID-19 Tweets Classification Based on a Hybrid Word Embedding Method. *Big Data Cogn. Comput.* **2022**, *6*, 58. <https://doi.org/10.3390/bdcc6020058>.
7. Parikh, S.; Davoudi, A.; Yu, S.; Giraldo, C.; Schriver, E.; Mowery, D. Lexicon Development for COVID-19-related Concepts Using Open-source Word Embedding Sources: An Intrinsic and Extrinsic Evaluation. *JMIR Med. Inform.* **2021**, *9*, e21679. <https://doi.org/10.2196/21679>.
8. Sciandra, A. COVID-19 Outbreak through Tweeters’ Words: Monitoring Italian Social Media Communication about COVID-19 with Text Mining and Word Embeddings. In Proceedings of the 2020 IEEE Symposium on Computers and Communications (ISCC), Rennes, France, 7–10 July 2020; pp. 1–6. <https://doi.org/10.1109/ISCC50000.2020.9219595>.

9. Levy, O.; Goldberg, Y.; Dagan, I. Improving Distributional Similarity with Lessons Learned from Word Embeddings. *Trans. Assoc. Comput. Linguist.* **2015**, *3*, 211–225. [https://doi.org/10.1162/tacl\\_a\\_00134](https://doi.org/10.1162/tacl_a_00134).
10. Firth, J.R. A Synopsis of Linguistic Theory 1930–1955. In *Studies in Linguistic Analysis*; Special Volume of the Philological Society; Blackwell: Oxford, UK, 1957; pp. 1–32.
11. Harris, Z.S. *Distributional Structure*; Routledge: New York, NY, USA, 1954.
12. Mikolov, T.; Corrado, G.; Chen, K.; Dean, J. Efficient Estimation of Word Representations in Vector Space. In Proceedings of the International Conference on Learning Representations, Scottsdale, AZ, USA, 2–4 May 2013; pp. 1–12.
13. Greenaway, K.H.; Kaloerinos, E.K.; Williams, L.A. Context is Everything (in Emotion Research). *Soc. Personal. Psychol. Compass* **2018**, *12*, 12393. <https://doi.org/10.1111/spc3.12393>.
14. Barrett, L.F.; Mesquita, B.; Smith, E.R. The context principle. In *The Mind in Context*; Mesquita, B., Barrett, L.F., Smith, E.R., Eds.; Guilford Press: New York, NY, USA, 2010; pp. 1–22.
15. Ledgerwood, A. Evaluations in their social context: Distance regulates consistency and context dependence. *Soc. Personal. Psychol. Compass* **2014**, *8*, 436–447.
16. Moskowitz, J.T.; Cheung, E.O.; Freedman, M.; Fernando, C.; Zhang, M.W.; Huffman, J.C.; Addington, E.L. Measuring positive emotion outcomes in positive psychology interventions: A literature review. *Emot. Rev.* **2020**, *13*, 60–73. <https://doi.org/10.1177/1754073920950811>.
17. Sun, R.; Balabanova, A.; Bajada, C.J.; Liu, Y.; Kriuchok, M.; Voolma, S.; Pavarini, G. Psychological wellbeing during the global COVID-19 outbreak. *PsyArXiv* **2020**. <https://doi.org/10.31234/osf.io/r7xaz>.
18. Welles, B.F.; González-Bailón, S. *The Oxford Handbook of Networked Communication*; Oxford University Press: Oxford, UK, 2020. ISBN 100190460512.
19. Basile, V.; Cauteruccio, F.; Terracina, G. How Dramatic Events Can Affect Emotionality in Social Posting: The Impact of COVID-19 on Reddit. *Future Internet* **2021**, *13*, 29. <https://doi.org/10.3390/fi13020029>.
20. Subreddit Stats. 2023. Available online: <https://subredditstats.com/> (accessed on 19 March 2023).
21. Subreddit Lists. 2023. Available online: <https://redditlist.com/> (accessed on 19 March 2023).
22. Coronavirus Subreddit. Available online: <https://www.reddit.com/r/Coronavirus/> (accessed on 19 March 2023).
23. Reddiquette: An Informal Expression of the Values of Many Redditors, as Written by Redditors Themselves. Available online: <https://www.reddithelp.com/hc/en-us/articles/205926439> (accessed on 19 March 2023).
24. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed Representations of Words and Phrases and their Compositionality. In Proceedings of the NIPS'13: 26th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–8 December 2013; Volume 2, pp. 3111–3119.
25. Wu, G.; Feder, A.; Cohen, H.; Kim, J.J.; Calderon, S.; Charney, D.S.; Mathé, A.A. Understanding resilience. *Front. Behav. Neurosci.* **2013**, *7*, 10. <https://doi.org/10.3389/fnbeh.2013.00010>.
26. Rutter, M. Resilience as a dynamic concept. *Dev. Psychopathol.* **2012**, *24*, 335–344. <https://doi.org/10.1017/S0954579412000028>.
27. Newman, R. APA's resilience initiative. *Prof. Psychol. Res. Pract.* **2005**, *36*, 227–229.
28. Vella, S.; Pai, N. A theoretical review of psychological resilience: Defining resilience and resilience research over the decades. *Arch. Med. Health Sci.* **2019**, *7*, 233–239.
29. Tariq, H. Measuring Community Disaster Resilience at local levels: An adaptable Resilience Framework. *Int. J. Disaster Risk Reduct.* **2021**, *62*, 102358. <https://doi.org/10.1016/j.ijdrr.2021.102358>.
30. Israelashvili, J. More Positive Emotions During the COVID-19 Pandemic Are Associated with Better Resilience, Especially for Those Experiencing More Negative Emotions. *Front. Psychol.* **2021**, *12*, 648112. <https://doi.org/10.3389/fpsyg.2021.648112>.
31. Zhang, Y.; Wang, X.; Lai, S.; He, S.; Liu, K.; Zhao, J.; Lv, X. Ontology Matching with Word Embeddings. In *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data*; Sun, M., Liu, Y., Zhao, J., Eds.; NLP-NABD CCL 2014, Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2014; Volume 8801. [https://doi.org/10.1007/978-3-319-12277-9\\_4](https://doi.org/10.1007/978-3-319-12277-9_4).
32. Battig, W.F.; Montague, W.E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut norms. *J. Exp. Psychol.* **1969**, *80*, 1–46. <https://doi.org/10.1037/h0027577>.
33. Van Overschelde, J.; Rawson, K.; Dunlosky, J. Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *J. Mem. Lang.* **2004**, *50*, 289–335. <https://doi.org/10.1016/j.jml.2003.10.003>.
34. Rajput, N.K.; Grover, B.A.; Rath, V.K. Word frequency and sentiment analysis of twitter messages during coronavirus pandemic. *arXiv* **2020**. arXiv:2004.03925.
35. Samuel, J.; Ali, G.; Rahman, M.; Esawi, E.; Samuel, Y. COVID-19 public sentiment insights and machine learning for tweets classification. *Information* **2020**, *11*, 314.
36. Aljameel, S.S.; Alabbad, D.A.; Alzahrani, N.A.; Alqarni, S.M.; Alamoudi, F.A.; Babili, L.M.; Aljaafary, S.K.; Alshamrani, F.M. A sentiment analysis approach to predict an individual's awareness of the precautionary procedures to prevent COVID-19 outbreaks in Saudi Arabia. *Int. J. Environ. Res. Public Health* **2021**, *18*, 218.
37. Muthusami, R.; Bharathi, A.; Saritha, K. COVID-19 outbreak: Tweet based analysis and visualization towards the influence of coronavirus in the world. *Gedrag Organ. Rev.* **2020**, *33*, 8–9.
38. Jalil, Z.; Abbasi, A.; Javed, A.R.; Badruddin Khan, M.; Abul Hasanat, M.H.; Malik, K.M.; Saudagar, A.K.J. COVID-19 Related Sentiment Analysis Using State-of-the-Art Machine Learning and Deep Learning Techniques. *Front. Public Health* **2022**, *9*, 812735. <https://doi.org/10.3389/fpubh.2021.812735>.

39. Rustam, F.; Khalid, M.; Aslam, W.; Rupapara, V.; Mehmood, A.; Choi, G.S. A performance comparison of supervised machine learning models for COVID-19 tweets sentiment analysis. *PLoS ONE* **2021**, *16*, e0245909.
40. Dangi, D.; Dixit, D.K.; Bhagat, A. Sentiment analysis of COVID-19 social media data through machine learning. *Multimed. Tools Appl.* **2022**, *81*, 42261–42283. <https://doi.org/10.1007/s11042-022-13492-w>.
41. Rahman, M.M.; Islam, M.N. Exploring the Performance of Ensemble Machine Learning Classifiers for Sentiment Analysis of COVID-19 Tweets. In *Sentimental Analysis and Deep Learning*; Shakya, S., Balas, V.E., Kamolphiwong, S., Du, K.L., Eds.; Advances in Intelligent Systems and Computing; Springer: Singapore, 2022; Volume 1408. [https://doi.org/10.1007/978-981-16-5157-1\\_30](https://doi.org/10.1007/978-981-16-5157-1_30).
42. Es-Sabery, F.; Es-Sabery, K.; Qadir, J.; Sainz-De-Abajo, B.; Hair, A.; Garcia-Zapirain, B.; De la Torre-Diez, I. A MapReduce opinion mining for COVID-19-related tweets classification using enhanced ID3 decision tree classifier. *IEEE Access* **2021**, *9*, 58706–58739. <https://doi.org/10.1109/ACCESS.2021.3073215>.
43. Basiri, M.E.; Nemati, S.; Abdar, M.; Asadi, S.; Acharrya, U.R. A novel fusion-based deep learning model for sentiment analysis of COVID-19 tweets. *Knowl.-Based Syst.* **2021**, *228*, 107242. <https://doi.org/10.1016/j.knosys.2021.107242>.
44. Ibrahim, F.A.; Hassaballah, M.; Ali, A.A.; Nam, Y.; Ibrahim, A.I. COVID19 outbreak: A hierarchical framework for user sentiment analysis. *Comput. Mater. Contin.* **2022**, *70*, 2507–2524. <https://doi.org/10.32604/cmc.2022.018131>.
45. Bonifazi, G.; Breve, B.; Cirillo, S.; Corradini, E.; Virgili, L. Investigating the COVID-19 vaccine discussions on Twitter through a multilayer network-based approach. *Inf. Process Manag.* **2022**, *59*, 103095. <https://doi.org/10.1016/j.ipm.2022.103095>.
46. Naseem, U.; Razzak, I.; Khushi, M.; Eklund, P.W.; Kim, J. Covidsenti: A large-scale benchmark Twitter data set for COVID-19 sentiment analysis. *IEEE Trans. Comput. Soc. Syst.* **2021**, *8*, 1003–1015.
47. Yan, C.; Law, M.; Nguyen, S.; Cheung, J.; Kong, J. Comparing public sentiment toward COVID-19 vaccines across Canadian cities: Analysis of comments on reddit. *J. Med. Internet Res.* **2021**, *23*, e32685. <https://doi.org/10.2196/32685>.
48. Jelodar, H.; Wang, Y.; Orji, R.; Huang, S. Deep Sentiment classification and topic discovery on novel coronavirus or COVID-19 online discussions: NLP using LSTM recurrent neural network approach. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 2733–2742.
49. Lai, D.; Wang, D.; Calvano, J.; Raja, A.S.; He, S. Addressing immediate public coronavirus (COVID-19) concerns through social media: Utilizing Reddit's AMA as a framework for public engagement with science. *PLoS ONE* **2020**, *15*, e0240326.
50. Pal, R.; Chopra, H.; Awasthi, R.; Bandhey, H.; Nagori, A.; Sethi, T. Predicting Emerging Themes in Rapidly Expanding COVID-19 Literature with Unsupervised Word Embeddings and Machine Learning: Evidence-Based Study. *J. Med. Internet Res.* **2022**, *24*, e34067. <https://doi.org/10.2196/34067>.
51. Jha, R.A.; Ananthanarayana, V.S. Gaining Actionable Insights in COVID-19 Dataset Using Word Embeddings. In *Pattern Recognition and Data Analysis with Applications*. In *Lecture Notes in Electrical Engineering*; Gupta, D., Goswami, R.S., Banerjee, S., Tanveer, M., Pachori, R.B., Eds.; Springer: Singapore, 2022; Volume 888. [https://doi.org/10.1007/978-981-19-1520-8\\_37](https://doi.org/10.1007/978-981-19-1520-8_37).
52. Batzdorfer, V.; Steinmetz, H.; Biella, M.; Alizadeh, M. Conspiracy theories on Twitter: Emerging motifs and temporal dynamics during the COVID-19 pandemic. *Int. J. Data Sci. Anal.* **2022**, *13*, 315–333. <https://doi.org/10.1007/s41060-021-00298-6>.
53. Bhandari, A.; Kumar, V.; Thien Huong, P.; Thanh, D. Sentiment Analysis of COVID-19 Tweets: Leveraging Stacked Word Embedding Representation for Identifying Distinct Classes Within a Sentiment. In *Artificial Intelligence in Data and Big Data Processing*; Dang, N.H.T., Zhang, Y.D., Tavares, J.M.R.S., Chen, B.H., Eds.; ICABDE 2021, Lecture Notes on Data Engineering and Communications Technologies; Springer: Cham, Switzerland, 2022; Volume 124. [https://doi.org/10.1007/978-3-030-97610-1\\_27](https://doi.org/10.1007/978-3-030-97610-1_27).
54. Chan, A.Y.; Ting, C.; Chan, L.G.; Hildon, Z.J.L. “The emotions were like a roller-coaster”: A qualitative analysis of e-diary data on healthcare worker resilience and adaptation during the COVID-19 outbreak in Singapore. *Hum. Resour. Health* **2022**, *20*, 60. <https://doi.org/10.1186/s12960-022-00756-7>.
55. Pushshift Reddit API Documentation. Available online: <https://github.com/pushshift/api> (accessed on 19 March 2023).
56. Lama, Y.; Hu, D.; Jamison, A.; Quinn, S.C.; Broniatowski, D.A. Characterizing Trends in Human Papillomavirus Vaccine Discourse on Reddit (2007–2015): An Observational Study. *JMIR Public Health Surveill.* **2019**, *5*, e12480. <https://doi.org/10.2196/12480>.
57. PushshiftR: An R Package for Connection to the Pushshift.io API. Available online: <https://github.com/dashstander/pushshiftR> (accessed on 19 March 2023).
58. Benoit, K.; Watanabe, K.; Wang, H.; Nulty, P.; Obeng, A.; Müller, S.; Matsuo, A. quanteda: An R package for the quantitative analysis of textual data. *J. Open Source Softw.* **2018**, *3*, 774. <https://doi.org/10.21105/joss.00774>. <https://quanteda.io>.
59. Silge, J.; Robinson, D. tidytext: Text Mining and Analysis Using Tidy Data Principles in R. *J. Open Source Softw.* **2016**, *1*, 37. <https://doi.org/10.21105/joss.00037>.
60. dplyr: A Grammar of Data Manipulation. Available online: <https://cran.r-project.org/web/packages/dplyr/index.html> (accessed on 19 March 2023).
61. ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics. Available online: <https://cran.r-project.org/web/packages/ggplot2/> (accessed on 19 March 2023).
62. broom: Convert Statistical Objects into Tidy Tibbles. Available online: <https://cran.r-project.org/web/packages/broom/index.html> (accessed on 19 March 2023).
63. wordVectors: An R Package for Building and Exploring Word Embedding Models. Available online: <https://github.com/bmschmidt/wordVectors> (accessed on 19 March 2023).

64. Rousseeuw, P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
65. Barter, R.L.; Yu, B. Superheat: An R package for creating beautiful and extendable heatmaps for visualizing complex data. *J. Comput. Graph. Stat.* **2018**, *27*, 910–922. <https://doi.org/10.1080/10618600.2018.1473780>.
66. Garcia-Rudolph, A.; Sauri, J.; Cegarra, B.; Bernabeu Guitart, M. Discovering the Context of People with Disabilities: Semantic Categorization Test and Environmental Factors Mapping of Word Embeddings from Reddit. *JMIR Med. Inform.* **2020**, *8*, e17903. Published 2020 Nov 20. <https://doi.org/10.2196/17903>.
67. The Official Website of the Government of Canada. Available online: <https://www.btb.termiumpplus.gc.ca/publications/covid19-eng.html> (accessed on 19 March 2023).
68. Mahendiratta, S.; Bansal, S.; Sarma, P.; Kumar, H.; Choudhary, G.; Kumar, S.; Prakash, A.; Sehgal, R.; Medhi, B. Stem cell therapy in COVID-19: Pooled evidence from SARS-CoV-2, SARS-CoV, MERS-CoV and ARDS: A systematic review. *Biomed. Pharma-cother.* **2021**, *137*, 111300. <https://doi.org/10.1016/j.biopha.2021.111300>.
69. Lazzeri, C.; Bonizzoli, M.; Batacchi, S.; Di Valvasone, S.; Chiostrì, M.; Peris, A. The prognostic role of hyperglycemia and glucose variability in covid-related acute respiratory distress Syndrome. *Diabetes Res. Clin. Pract.* **2021**, *175*, 108789. <https://doi.org/10.1016/j.diabres.2021.108789>.
70. Chilosi, M.; Poletti, V.; Ravaglia, C.; Rossi, G.; Dubini, A.; Piciucchi, S.; Pedica, F.; Bronte, V.; Pizzolo, G.; Martignoni, G.; et al. The pathogenic role of epithelial and endothelial cells in early-phase COVID-19 pneumonia: Victims and partners in crime. *Mod. Pathol.* **2021**, *34*, 1444–1455. <https://doi.org/10.1038/s41379-021-00808-8>.
71. Helms, J.; Severac, F.; Merdji, H.; Schenck, M.; Clere-Jehl, R.; Baldacini, M.; Ohana, M.; Grunebaum, L.; Castelain, V.; Anglés-Cano, E.; et al. Higher anticoagulation targets and risk of thrombotic events in severe COVID-19 patients: Bi-center cohort study. *Ann. Intensive Care* **2021**, *11*, 14. <https://doi.org/10.1186/s13613-021-00809-5>.
72. Chang, C.C.; Yang, M.H.; Chang, S.M.; Hsieh, Y.J.; Lee, C.H.; Chen, Y.A.; Yuan, C.H.; Chen, Y.L.; Ho, S.Y.; Tyan, Y.C. Clinical significance of olfactory dysfunction in patients of COVID-19. *J. Chin. Med. Assoc.* **2021**, *84*, 682–689. <https://doi.org/10.1097/JCMA.0000000000000560>.
73. Rethinavel, H.S.; Ravichandran, S.; Radhakrishnan, R.K.; Kandasamy, M. COVID-19 and Parkinson's disease: Defects in neurogenesis as the potential cause of olfactory system impairments and anosmia. *J. Chem. Neuroanat.* **2021**, *115*, 101965. <https://doi.org/10.1016/j.jchemneu.2021.101965>. Epub ahead of print.
74. Buchheit, K.; Bensko, J.C.; Lewis, E.; Gakpo, D.; Laidlaw, T.M. The importance of timely diagnosis of aspirin-exacerbated respiratory disease for patient health and safety. *World J. Otorhinolaryngol. Head Neck Surg.* **2020**, *6*, 203–206. <https://doi.org/10.1016/j.wjorl.2020.07.003>.
75. Vandergaast, R.; Carey, T.; Reiter, S.; Lathrum, C.; Lech, P.; Gnanadurai, C.; Haselton, M.; Buehler, J.; Narjari, R.; Schnebeck, L.; et al. IMMUNO-COV v2.0: Development and Validation of a High-Throughput Clinical Assay for Measuring SARS-CoV-2-Neutralizing Antibody Titers. *mSphere* **2021**, *6*, e0017021. <https://doi.org/10.1128/mSphere.00170-21>.
76. Baum, A.; Kyratsous, C.A. SARS-CoV-2 spike therapeutic antibodies in the age of variants. *J. Exp. Med.* **2021**, *218*, e20210198. <https://doi.org/10.1084/jem.20210198>.
77. Calitri, C.; Fantone, F.; Benetti, S.; Lupica, M.M.; Ignaccolo, M.G.; Banino, E.; Viano, A.; Pace, M.; Castella, A.; Gaido, F.; et al. Long-term clinical and serological follow-up of paediatric patients infected by SARS-CoV-2. *Infez Med.* **2021**, *29*, 216–223.
78. Kutzler, H.L.; Kuzaro, H.A.; Serrano, O.K.; Feingold, A.; Morgan, G.; Cheema, F. Initial Experience of Bamlanivimab Monotherapy Use in Solid Organ Transplant Recipients. *Transpl. Infect. Dis.* **2021**, *23*, e13662. <https://doi.org/10.1111/tid.13662>.
79. Wadaa-Allah, A.; Emhamed, M.S.; Sadeq, M.A.; Ben Hadj Dahman, N.; Ullah, I.; Farrag, N.S.; Negida, A. Efficacy of the current investigational drugs for the treatment of COVID-19: A scoping review. *Ann. Med.* **2021**, *53*, 318–334. <https://doi.org/10.1080/07853890.2021.1875500>.
80. Hu, Y.; Meng, X.; Zhang, F.; Xiang, Y.; Wang, J. The in vitro antiviral activity of lactoferrin against common human coronaviruses and SARS-CoV-2 is mediated by targeting the heparan sulfate co-receptor. *Emerg. Microbes Infect.* **2021**, *10*, 317–330. <https://doi.org/10.1080/22221751.2021.1888660>.
81. Vergori, A.; Lorenzini, P.; Cozzi-Lepri, A.; Donno, D.R.; Gualano, G.; Nicastri, E.; Iacomì, F.; Marchioni, L.; Campioni, P.; Schinà, V.; et al. Prophylactic heparin and risk of orotracheal intubation or death in patients with mild or moderate COVID-19 pneumonia. *Sci. Rep.* **2021**, *11*, 11334. <https://doi.org/10.1038/s41598-021-90713-6>.
82. Li, C.; Luo, F.; Liu, C.; Xiong, N.; Xu, Z.; Zhang, W.; Yang, M.; Wang, Y.; Liu, D.; Yu, C.; et al. Effect of a genetically engineered interferon-alpha versus traditional interferon-alpha in the treatment of moderate-to-severe COVID-19: A randomised clinical trial. *Ann. Med.* **2021**, *53*, 391–401. <https://doi.org/10.1080/07853890.2021.1890329>.
83. Daoud, S.; Alabed, S.J.; Dahabiyeh, L.A. Identification of potential COVID-19 main protease inhibitors using structure-based pharmacophore approach, molecular docking and repurposing studies. *Acta Pharm.* **2021**, *71*, 163–174. <https://doi.org/10.2478/acph-2021-0016>.
84. Liu, Y.; Cooper, C.L.; Tarba, S.Y. Resilience, wellbeing and HRM: A multidisciplinary perspective. *Int. J. Hum. Resour. Manag.* **2019**, *30*, 1227–1238. <https://doi.org/10.1080/09585192.2019.1565370>.
85. Brog, N.A.; Hegy, J.K.; Berger, T.; Znoj, H. An internet-based self-help intervention for people with psychological distress due to COVID-19: Study protocol for a randomized controlled trial. *Trials* **2021**, *22*, 171. <https://doi.org/10.1186/s13063-021-05089-9>.

86. Park, C.L.; Finkelstein-Fox, L.; Russell, B.S.; Fendrich, M.; Hutchison, M.; Becker, J. Psychological resilience early in the COVID-19 pandemic: Stressors, resources, and coping strategies in a national sample of Americans. *Am. Psychol.* **2021**, *76*, 715–728. <https://doi.org/10.1037/amp0000813>.
87. Ameis, S.H.; Lai, M.C.; Mulsant, B.H.; Szatmari, P. Coping, fostering resilience, and driving care innovation for autistic people and their families during the COVID-19 pandemic and beyond. *Mol. Autism* **2020**, *11*, 61. <https://doi.org/10.1186/s13229-020-00365-y>.
88. Tafoya, S.A.; Aldrete-Cortez, V.; Ortiz, S.; Fouilloux, C.; Flores, F.; Monterrosas, A.M. Resilience, sleep quality and morningness as mediators of vulnerability to depression in medical students with sleep pattern alterations. *Chronobiol. Int.* **2019**, *36*, 381–391. <https://doi.org/10.1080/07420528.2018.1552290>.
89. Ungar, M.; Ghazinour, M.; Richter, J. Annual Research Review: What is resilience within the social ecology of human development? *J. Child Psychol. Psychiatry* **2013**, *54*, 348–366. <https://doi.org/10.1111/jcpp.12025>.
90. Yang, C.; Zhou, Y.; Xia, M. How Resilience Promotes Mental Health of Patients with DSM-5 Substance Use Disorder? The Mediation Roles of Positive Affect, Self-Esteem, and Perceived Social Support. *Front. Psychiatry* **2020**, *11*, 588968. <https://doi.org/10.3389/fpsy.2020.588968>.
91. Sterina, E.; Hermida, A.P.; Gerberi, D.J.; Lapid, M.I. Emotional Resilience of Older Adults during COVID-19: A Systematic Review of Studies of Stress and Well-Being. *Clin. Gerontol.* **2021**, *45*, 4–19. <https://doi.org/10.1080/07317115.2021.1928355>.
92. Buchman, A.S.; Yu, L.; Oveisgharan, S.; Petyuk, V.A.; Tasaki, S.; Gaiteri, C.; Wilson, R.S.; Grodstein, F.; Schneider, J.A.; Klein, H.U.; et al. Cortical proteins may provide motor resilience in older adults. *Sci. Rep.* **2021**, *11*, 11311. <https://doi.org/10.1038/s41598-021-90859-3>.
93. Koerner, S.S.; Shirai, Y. Latina/o Family Caregivers' Reactions to Limited Help from Relatives: From Frustration to Resilience. *J. Fam. Nurs.* **2019**, *25*, 590–609. <https://doi.org/10.1177/1074840719867706>.
94. Jané-Llopis, E.; Anderson, P.; Segura, L.; Zabaleta, E.; Muñoz, R.; Ruiz, G.; Rehm, J.; Cabezas, C.; Colom, J. Mental ill-health during COVID-19 confinement. *BMC Psychiatry* **2021**, *21*, 194. <https://doi.org/10.1186/s12888-021-03191-5>.
95. Brant-Birioukov, K. COVID-19 and In(di)genuity: Lessons from Indigenous resilience, adaptation, and innovation in times of crisis. *Prospects* **2021**, *51*, 247–259. <https://doi.org/10.1007/s11125-021-09557-7>.
96. Catungal, J.P. Essential workers and the cultural politics of appreciation: Sonic, visual and mediated geographies of public gratitude in the time of COVID-19. *Cult. Geogr.* **2021**, *28*, 403–408. <https://doi.org/10.1177/1474474020978483>.
97. Elias, A.; Ben, J.; Mansouri, F.; Paradies, Y. Racism and nationalism during and beyond the COVID-19 pandemic. *Ethn. Racial Stud.* **2021**, *44*, 783–793. <https://doi.org/10.1080/01419870.2020.1851382>.
98. Croucher, S.M.; Nguyen, T.; Rahmani, D. Prejudice toward Asian Americans in the Covid-19 Pandemic: The Effects of Social Media use in the United States. *Front. Commun.* **2020**, *5*, 39. <https://doi.org/10.3389/fcomm.2020.00039>.
99. Devakumar, D.; Shannon, G.; Bhopal, S.S.; Abubakar, I. Racism and Discrimination in COVID-19 Responses. *Lancet* **2020**, *395*, 1194. [https://doi.org/10.1016/S0140-6736\(20\)30792-3](https://doi.org/10.1016/S0140-6736(20)30792-3).
100. Selman, L.E.; Chamberlain, C.; Sowden, R.; Chao, D.; Selman, D.; Taubert, M.; Braude, P. Sadness, despair and anger when a patient dies alone from COVID-19: A thematic content analysis of Twitter data from bereaved family members and friends. *Palliat. Med.* **2021**, *35*, 1267–1276. <https://doi.org/10.1177/02692163211017026>.
101. Gozzi, N.; Tizzani, M.; Starnini, M.; Ciulla, F.; Paolotti, D.; Panisson, A.; Perra, N. Collective response to media coverage of the COVID-19 pandemic on Reddit and Wikipedia: Mixed-methods analysis. *J. Med. Internet Res.* **2020**, *22*, e21597.
102. Stetten, N.E.; LeBeau, K.; Aguirre, M.A.; Vogt, A.B.; Quintana, J.R.; Jennings, A.R.; Hart, M. Analyzing the Communication Interchange of Individuals with Disabilities Utilizing Facebook, Discussion Forums, and Chat Rooms: Qualitative Content Analysis of Online Disabilities Support Groups. *JMIR Rehabil. Assist. Technol.* **2019**, *6*, e12667.
103. Li, X.; Zhang, J.; Du, Y.; Zhu, J.; Fan, Y.; Chen, X. A Novel Deep Learning-based Sentiment Analysis Method Enhanced with Emojis in Microblog Social Networks. *Enterp. Inf. Syst.* **2022**, *17*, 2037160.
104. Corradini, E.; Nocera, A.; Ursino, D.; Virgili, L. Investigating the phenomenon of NSFW posts in Reddit. *Inf. Sci.* **2021**, *566*, 140–164.
105. Padilla, J.; Kavak, H.; Lynch, C.; Gore, R.; Diallo, S. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. *PLoS ONE* **2018**, *13*, e0198857.
106. Gore, R.; Diallo, S.; Padilla, J. You Are What You Tweet: Connecting the Geographic Variation in America's Obesity Rate to Twitter Content. *PLoS ONE* **2015**, *10*, e0133505.
107. Reddit's 2020 Year in Review. Available online: <https://redditblog.com/2020/12/08/reddits-2020-year-in-review/> (accessed on 20 March 2023).

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.