

Article

The Public Health Contribution of Sentiment Analysis of Monkeypox Tweets to Detect Polarities Using the CNN-LSTM Model

Orlando Iparraguirre-Villanueva ¹, Aldo Alvarez-Risco ² , Jose Luis Herrera Salazar ³, Saul Beltozar-Clemente ⁴, Joselyn Zapata-Paulini ⁵, Jaime A. Yáñez ^{6,*}  and Michael Cabanillas-Carbonell ^{7,*} 

¹ Facultad de Ingeniería y Arquitectura, Universidad Autónoma del Perú, Lima 15842, Peru

² Carrera de Negocios Internacionales Facultad de Ciencias Empresariales y Económicas, Universidad de Lima, Lima 15023, Peru

³ Facultad de Ingeniería, Ciencias y Administración, Universidad Autónoma de Ica, Chincha Alta 11701, Peru

⁴ Dirección de Cursos Básicos, Universidad Científica del Sur, Lima 15067, Peru

⁵ Escuela de Posgrado, Universidad Continental, Lima 12000, Peru

⁶ Vicerrectorado de Investigación, Universidad Norbert Wiener, Lima 15046, Peru

⁷ Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad Politécnica de Madrid, 28040 Madrid, Spain

* Correspondence: jaime.yanez@uwiener.edu.pe (J.A.Y.); mcabanillas@ieee.org (M.C.-C.)

Abstract: Monkeypox is a rare disease caused by the monkeypox virus. This disease was considered eradicated in 1980 and was believed to affect rodents and not humans. However, recent years have seen a massive outbreak of monkeypox in humans, setting off worldwide alerts from health agencies. As of September 2022, the number of confirmed cases in Peru had reached 1964. Although most monkeypox patients have been discharged, we cannot neglect the monitoring of the population with respect to the monkeypox virus. Lately, the population has started to express their feelings and opinions through social media, specifically Twitter, as it is the most used social medium and is an ideal space to gather what people think about the monkeypox virus. The information imparted through this medium can be in different formats, such as text, videos, images, audio, etc. The objective of this work is to analyze the positive, negative, and neutral feelings of people who publish their opinions on Twitter with the hashtag #Monkeypox. To find out what people think about this disease, a hybrid-based model architecture built on CNN and LSTM was used to determine the prediction accuracy. The prediction result obtained from the total monkeypox data was 83% accurate. Other performance metrics were also used to evaluate the model, such as specificity, recall level, and F1 score, representing 99%, 85%, and 88%, respectively. The results also showed the polarity of feelings through the CNN-LSTM confusion matrix, where 45.42% of people expressed neither positive nor negative opinions, while 19.45% expressed negative and fearful feelings about this infectious disease. The results of this work contribute to raising public awareness about the monkeypox virus.

Keywords: monkeypox; sentiment; tweets; CNN; LSTM



Citation: Iparraguirre-Villanueva, O.; Alvarez-Risco, A.; Herrera Salazar, J.L.; Beltozar-Clemente, S.; Zapata-Paulini, J.; Yáñez, J.A.; Cabanillas-Carbonell, M. The Public Health Contribution of Sentiment Analysis of Monkeypox Tweets to Detect Polarities Using the CNN-LSTM Model. *Vaccines* **2023**, *11*, 312. <https://doi.org/10.3390/vaccines11020312>

Academic Editor: Giuseppe La Torre

Received: 10 December 2022

Revised: 19 January 2023

Accepted: 20 January 2023

Published: 31 January 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

The monkeypox virus was first discovered in 1958 in colonies of monkeys being bred for research [1]. It was in the Democratic Republic of Congo that the first human case of monkeypox virus was reported in 1970 [2]. Since then, vaccines against the virus have been developed. In 1980, the monkeypox virus was declared eradicated and vaccination of the population was stopped [3,4]. While it is true that most patients can recover, the mortality rate during a monkeypox outbreak has traditionally ranged from 1% to 10% [5]. Monkeypox is a contagious disease that initially affected countries in Africa; however, it has recently spread to virtually every city in the world. Although the World Health Organization (WHO, Geneva, Switzerland) does not recognize it as a pandemic, some

experts believe it should be treated as such [6]. For example, the countries with the highest number of infections from May to 28 August are the United States with 403,937, Spain with 245,977, the United Kingdom with 159,815, Germany with 155,315, and France with 115,187, among others [7], as shown in Figure 1.

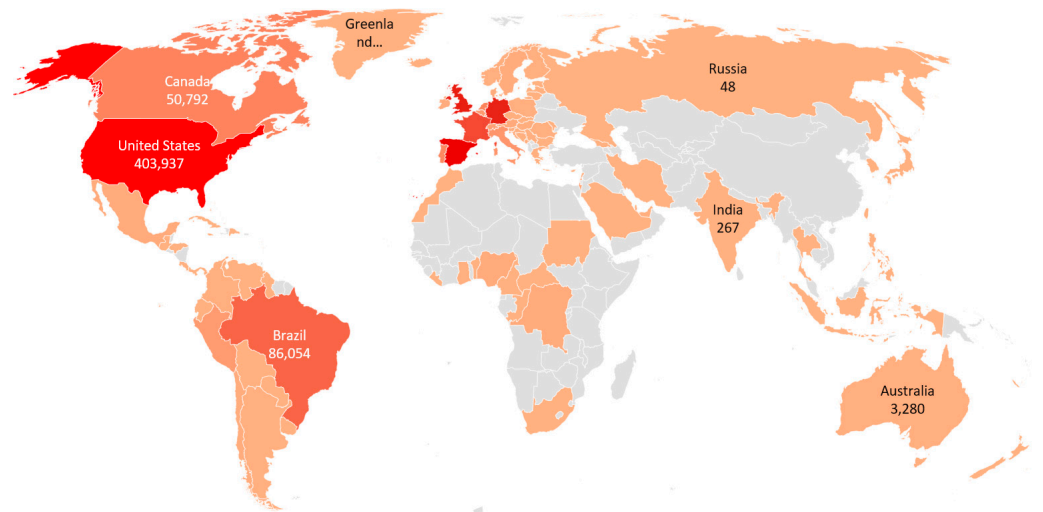


Figure 1. New and cumulative cases of monkeypox worldwide from May through 28 August 2022.

Currently, specialists in social data processing are making strenuous efforts to understand the human condition and the psychology of the human condition [8]. Twitter is considered a source of large volumes of data used in the field of behavioral sciences and human psychology, with the purpose of analyzing and knowing the feelings of Internet users [9]. With data processing, it is possible to automatically classify people's opinions regarding a specific topic into three categories: positive, negative, or neutral. In this work, a hybrid model was used to analyze and quantify the sentiments of Twitter users [10]. The main objective of this work was to analyze the positive, negative, and neutral sentiments posted on Twitter with the hashtag #ViruelaDelMono, through natural language processing (NLP) and machine learning algorithms. Twitter has become an important source of information dissemination through the Internet. The real-time, clear, and accurate communication of current events motivates users to register their opinions and reactions [11]. Twitter is the platform where users share their reactions, feelings, and opinions related to epidemics such as COVID-19, monkeypox, and other epidemics [12,13]. However, to collect all this valuable information provided by Twitter, the use of algorithms based on machine learning (ML) is required, due to a large number of words and contextual phrases that this represents for its processing [14,15]. Sentiment analysis is an NLP technique used to determine positive, negative, and neutral data. It is usually performed on textual data to monitor the opinions of Twitter users [16].

WHO/PAHO, academics, and scientists have published research papers on the epidemiological situation of the monkeypox virus [17,18]. For example, ref. [19] developed an extensible open-source tool that provides a comprehensive approach to assessing the accuracy of monkeypox virus-related claims in order to combat misinformation in digital media. Similarly, ref. [20] presented a regression model based on a CNN capable of handling cells. The algorithm of this model requires a set of training images generating output structures, commonly referred to as proximity patches [21]; these proximity patches present values closer to the center of the cells [22]. On the other hand, Italy has been a very important population focus for studies related to the subject, considering that it was the country with the highest number of people infected by COVID-19 in a given period. In view of this, a model was developed using artificial neural networks to determine potential donors for treatment with COVID-19 convalescent plasma [23]. Along the same lines, two studies developed a model capable of analyzing the themes and sentiments associated with mis-

information about the COVID-19 vaccine in social networks [24,25], for which the latent Dirichlet allocation (LDA) machine learning model is used to identify the predominant themes in the misinformation of sentiment-analysis-related research [26]. Classification algorithms are those used for data mining and text mining. That is why the authors of [27] conducted a search on Twitter to collect the main tweets with the word “monkeypox”. The analysis showed that 60% of the tweets posted corresponded to misinformation, followed by information provided by public health officials (which represented 32%) and reports from media or journalists (which represented 8%). To arrive at these results, the information in the tweets was divided into medically correct information and information that was not medically correct. The data published on social networks are data that are worthless without processing [28]. This is why it is important to process these data with the appropriate models. Although there is a set of machine learning models for prediction, it is important to consider the accuracy rates of each one. Therefore, ref. [29] compared 13 pre-trained deep learning models for monkeypox virus detection, using precision, recall, F1 score, and accuracy as measures, with results of 85.44%, 85.47%, 85.40%, and 87.13%, respectively. The present research is organized as follows: Section 1 addresses the research problem, the importance of the study, the purpose, and the objective to be carried out. Section 2 describes the work methodology, the LSTM architecture, the process diagram, and the dataset to be trained. Section 3 presents the results obtained after training. Section 4 presents the discussion. Finally, Section 5 presents the conclusions of the work.

2. Material and Methods

This section describes the terminology corresponding to the CNN-LSTM model and the implementation of the work to analyze the positive, negative, and neutral sentiments posted on Twitter Peru with the hashtag #ViruelaDelMono (Spanish for monkeypox).

2.1. CNN-LSTM

The hybrid CNN-LSTM model provides more efficient data processing than the stand-alone models. The convolutional layers function as filters of the incoming data, extracting relevant information that is used as input for the new connected convolutional layer. LSTM is a subcategory of the recurrent neural network (RNN); thus, it works very similarly to RNN. This network can add or remove information that it considers relevant for processing. Compared to RNN cells, LSTM cells have an additional input, which is known as a state cell. This cell is the key to the processing of LSTM networks and can function as a conveyor belt to which irrelevant data that we do not want to remain in the network memory can be added or removed. For this work on sentiment categorization based on Twitter data, a hybrid architecture called CNN-LSTM is used to test the model (Figure 2).

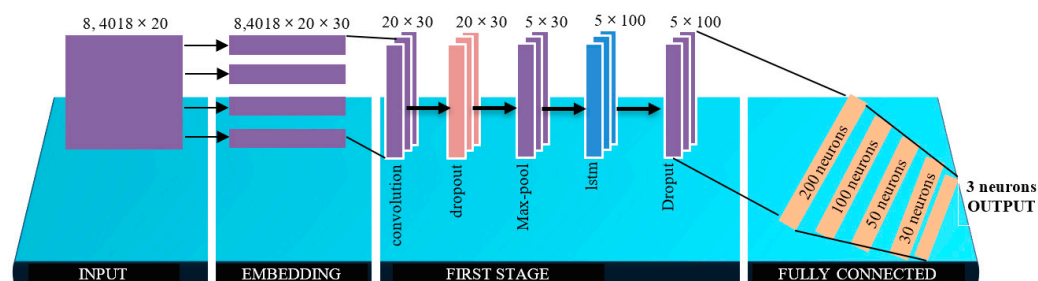


Figure 2. CNN-LSTM architecture for sentiment analysis.

In general, traditional RNNs seek to solve problems related to “memory loss”, as this generates unfavorable performance in time series and sequence problems. These models interact cyclically in the hidden layer, thus developing short-term memory, and can extract time series data. However, one of the limitations of RNNs is the loss gradient, which makes it difficult for the model to continue learning. Therefore, LSTMs address this problem by

retaining important information in memory cells and removing unimportant information, which results in better performance than a conventional RNN.

The LSTM network has 3 main gates: input, output, and forgetting. This scheme allows the network to maintain a controlled flow of information to determine which data to forget and which to remember, thus learning long-term dependencies. The input gate i_t , together with the second gate, regulates the volume of new memory states C_t at time t . The forgetting gate f determines whether information from the past should be erased or retained in the memory at time $t - 1$, while the output gate o determines which information can be used for the output of the memory cell. Equations (2)–(5) of the activities of an LSTM unit are described below.

$$i_t = \sigma(U_i x_t + w_i h_{t-1} + b_i) \tag{1}$$

$$f_t = \sigma(U_f x_t + w_f h_{t-1} + b_f) \tag{2}$$

$$c_t = \tanh(U_c x_t + w_c h_{t-1} + b_c) \tag{3}$$

$$c_t = g_t \cdot c_{t-1} + i_t \cdot c_t \tag{4}$$

$$O_t = \sigma(U_o x_t + w_o h_{t-1} + b_o) \tag{5}$$

where x_t represents the input, W^* and U^* represent the weight matrices, b^* represents the bias term vectors, σ represents the sigmoid function, and the operator \cdot represents the multiplication.

$$h_t = O_t \oplus \tanh(c_t) \tag{6}$$

In general, the CNN-LSTM model uses the CNN as an encoder for capturing input data characteristics in an LSTM. Figure 1 shows the architecture.

2.2. Understanding Data

To develop the case, we used a data set consisting of 84,018 tweets related to the monkeypox virus, which were obtained from Peru in September 2022. As shown in Figure 3, the process involved a series of steps, starting with the import of the dataset, followed by data processing, early cleaning, tokenization, exploratory analysis, classification, and finally the sentiment analysis of opinions. In the following sections, we will take a closer look at each of the points described above in more detail.

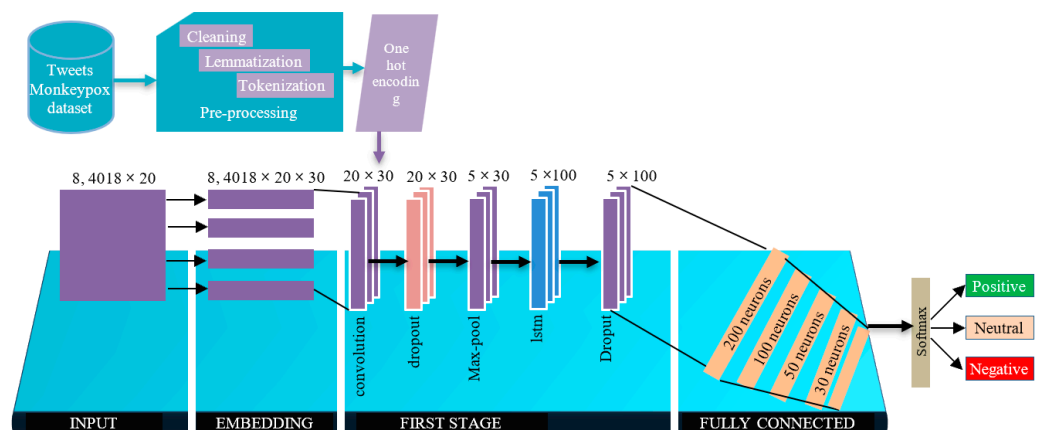


Figure 3. General sentiment classification and analysis process.

2.3. Data Cleaning and Processing

Figure 2 describes the general process of sentiment analysis and classification. However, before performing the actual training with tweets, preprocessing is an important step that must be performed first. This starts with the import of the dataset of real tweets.

For this work, we managed to extract 84,018 tweets with the hashtag #ViruelaDelMono corresponding to the month of September 2022 in Peru.

Therefore, to carry out the classification and prediction process, it was necessary first to carry out the cleaning process. This consisted in identifying, correcting, or eliminating incomplete, incorrect, or duplicate data that did not contribute to the model. This process was important to enable the model to make more accurate decisions regarding data analysis. This can include everything from checking the accuracy and completeness of the data to filling in missing data and converting the data to a new format. For the cleanup process, there was a set of libraries and tools that Python provides. The most relevant libraries in this work were the following:

1. Pandas: a data analysis library that provides a data structure and tools for cleaning and manipulating data. Panda allows users to eliminate duplicated data, complete missing values, and change column formats, among other cleaning operations.
2. Numpy: a numerical calculation library that provides functions for working with arrays and matrices. Numpy allows users to remove NaN (Not a Number) and perform math on the dataset.
3. OpenRefine: an open-source data-cleansing tool that allows users to quickly and easily clean, transform, and enhance large data sets. In this work, it was used to load data in formats such as CSV, Excel, JSON, and XML. Additionally, the regular expressions technique was released to search for patterns in the data set. This technique helps users find and replace specific patterns of data, that is, only the data that are useful to clean [30].

In this process, stemming was used to replace the words extracted from Twitter with the root word throughout the process. Then, we continued with tokenization, which consists of dividing the text of the sentences into words. Subsequently, a preliminary analysis was performed to understand and analyze the meaning of the words and how often they were used. This work used the programming language Python. For this preliminary analysis process, the DataFrame was used, in which the tweet information was stored. The next step was classification, for which the most commonly used method is the well-known bag of words, which consists of identifying the format for all the words provided by the corpus. Finally, the sentiment analysis was performed. For this analysis, it was necessary to have a dictionary of words in order to associate a feeling with each word.

The collected tweets included 84k tweets contained relevant information related to monkeypox. Within these 84k tweets, there were 90 attributes, so 90 attributes were processed with 84k tweets. Subsequently, lemmatization was applied. This process is based on the fact that the root of a word can be presented in various ways referring to the same thing. For this process, the UDpipe library was used, as shown in Table 1.

Table 1. Lemmatization of words.

Sentence_id	Sentence	Token_id	Token	Lema
1	Ministerio	1	Ministerio	Ministerio
1	Salud	1	Salud	Salud
1	Informa	1	Informa	Informar
1	Personas	1	Personas	Persona
1	Dadas	1	Dadas	Dada
1	Alta	1	Alta	Alto
1	Detectado	1	Detectado	Detectado
1	Casos	1	Casos	Caso

Table 1. Cont.

Sentence_id	Sentence	Token_id	Token	Lema
1	vienen	1	vienen	Vienir
1	recibiendo	1	recibiendo	Recibir
1	médica	1	médica	Médicar
1	Viene	1	Viene	Venir
1	Informa	1	Informa	informar

The next step was to perform an exploratory analysis of the dataset in order to identify the most commonly used words in the corpus, using the bag of words model. This method is used to present words ignoring their order. In this model, a document is a bag of words, which allows for dictionary-based word modeling, where each bag contains a certain number of words. Figure 4 shows the word cloud, highlighting the most frequent words.



Figure 4. Most frequent words in the corpus.

Next, the Syuzhet software was used, which facilitates the extraction of text-derived topics using a variety of opinion dictionaries conveniently packaged for model consumption. This package implements the following dictionaries: Bing, Stanford, Afinn, and NRC. The NRC dictionary was used for this work since it is the only one in the Spanish language. This dictionary has more than 14 thousand words divided into categories of feelings. These categories include positive, negative, and angry emotions, such as anger, fear, disgust, rage, premonition, etc. This dictionary was used to group the words and to understand the categories into which the feelings are divided. Figure 5 shows the feelings that are most frequently used with respect to the monkeypox virus.

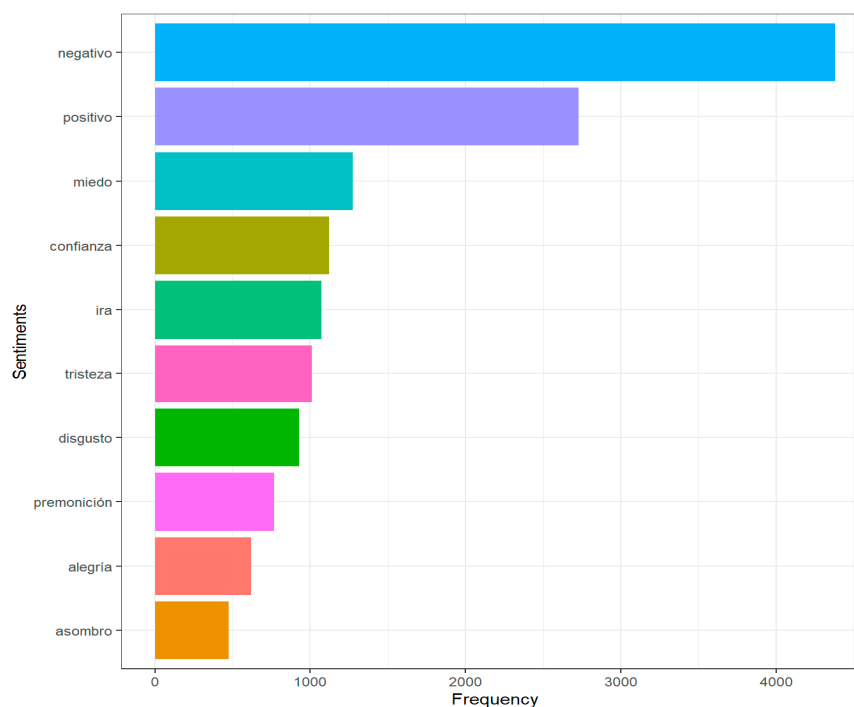


Figure 5. The most frequently used terms.

Figure 6 shows the words grouped by sentiment, which allowed us to analyze the set of words that are associated with a certain sentiment. For example, the feeling disgust is associated with the following words: enfermedad, contagio, epidemia, secuelas, muerte, and erupción. Along the same lines, the word “fear” is associated with the following words: caso, enfermedad, contagio, riesgo, gobierno, and evitar. In addition, negative sentiment is associated with words such as: virus, caso, enfermedad, riesgo, and evitar. Similarly, the feeling of sadness is associated with the following words: caso, enfermedad, evitar, grave, and emergencia. However, there are also words associated with positive sentiment, such as: ministerio, paciente, vacuna, información, and importante. It is important to look at the positive sentiment. Since there are a certain number of users who do not rate everything negatively, it means that they have confidence in the ministry and in the information provided by reliable media and rate the monkeypox virus vaccine positively.

Up to this stage, we had already cleaned and grouped the data in a document. In the next step, the TM package was used to employ the `removeSparseTerms()` function, allowing words that do not have many iterations or are not frequent to be removed. Additionally, in this phase, it was necessary to analyze the relationship between words; for this purpose, the function `finAssocs()` was used. This function calculates the relationship between words and assigns a score that is represented in the line thickness of the association between words. The tweets of the data set in this work were cleaned and organized in such a way that the operation of searching for a relationship between the terms was simple. It involved the use of a vector which could then be used to create the diagram with the `ggplot2()` library, as shown in Figure 7. For example, the root word `virueladelmono` is associated with `virueladelmono-salud`, `virueladelmono-contagios`, `virueladelmono-nuevos`, `virueladelmono-casos`, `virueladelmono-país`, and `virueladelmono-confirmados`. This is also followed by the following associations: `vacunación-registro`, `registro-virueladelmono`, `casos-ministerio`, `enfermedad-vacunación`, `vacunación-nacional`, `riesgo-vacunación`, `si-riesgo`, `virus-viruela símica`, and `entre`, among others.

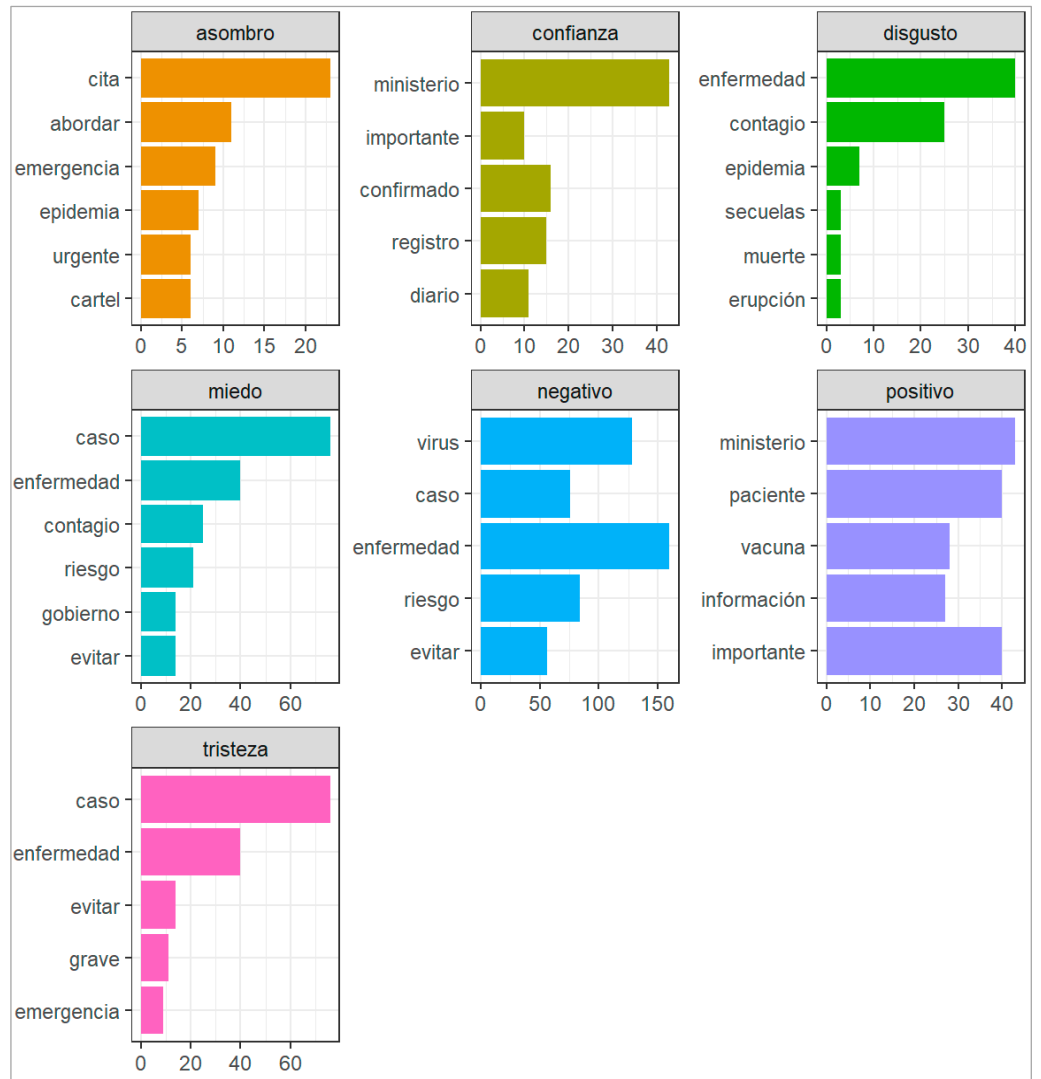


Figure 6. Words grouped by sentiment.

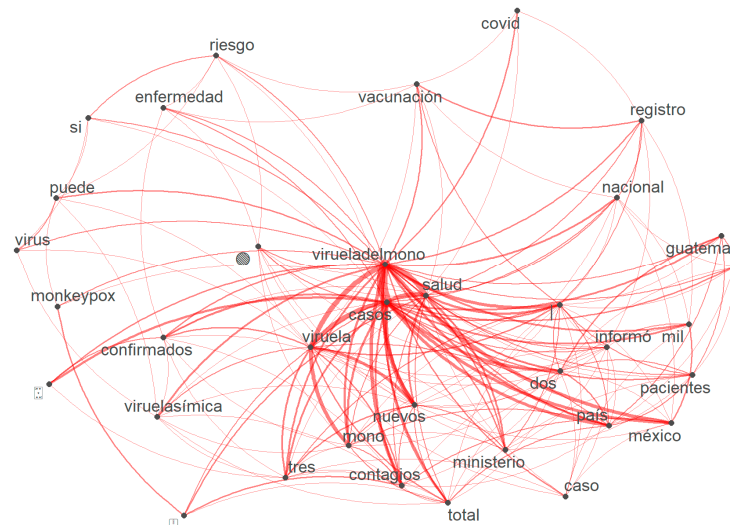


Figure 7. Word association and frequency.

3. Results

In this section, the performance of the CNN-LSTM model for sentiment prediction is presented. In this work, the CNN-LSTM model was used to optimize the classification performance and sentiment analysis. This hybrid model uses a set of intermediate layers which can be seen in Figures 1 and 2. For training, the model was applied at different times and with different amounts of data to obtain better model optimization; the training results are shown in Table 2. The results show that the model achieved good results in regard to the four variables: 83% was obtained in accuracy; a 99% rating was obtained in the specificity metric; 85% was obtained in the recovery metric; and, finally, 88% was obtained in the F1 score. Although it is true that 83% was obtained for precision, this is not the best precision metric; however, it can be considered optimal given that the greater the amount of data, the better the performance, and the precision of the model tends to improve.

Table 2. Model performance metrics of CNN-LSTM.

Model	Accuracy	Specificity	Recall	F1
CNN-LSTM1	0.83	0.98	0.89	0.88
CNN-LSTM2	0.83	0.99	0.88	0.89
CNN-LSTM3	0.85	0.98	0.97	0.87
CNN-LSTM4	0.82	0.99	0.76	0.88
CNN-LSTM5	0.79	0.99	0.77	0.89
CNN-LSTM6	0.81	0.98	0.86	0.88
CNN-LSTM7	0.84	0.98	0.85	0.87
CNN-LSTM8	0.86	0.99	0.79	0.88

Figure 8 shows the results after performing sentiment analysis on the data set related to the monkeypox. It can be seen from the results that the sentiments are very unequal, meaning that a large part of the users' opinions are neither positive nor negative, but neutral. Neutral sentiments represent 45.42% of the data, while negative opinions represent 19.45%. This indicates that the monkeypox virus does not represent generalized negative sentiments. However, 35.13% of opinions regarding the monkeypox virus are positive. Twitter users are probably no longer surprised by a new virus since it closely follows COVID-19. Additionally, users may feel more confident, since we are coming out of a very difficult experience and are now prepared.

Predicted/ Actuals	Positive	Negative	Neutral
Positive	35.13%	22.49%	42.38%
Negative	39.78%	19.45%	40.77%
Neutral	17.87%	30.71%	45.42%

Figure 8. Confusion matrix of CNN-LSTM: Analysis de sentiments.

The confusion matrix was used to represent the sentiment analysis and classification. Figure 8 demonstrates that the proposed method correctly classifies the sentiments of Twitter users regarding monkeypox.

4. Discussion

For this work, a total of 84,018 Twitter feeds were collected to be analyzed and classified with the purpose of determining the feelings of Twitter users regarding the impact of the monkeypox virus. First, people's reaction was expected to be one of the concerns, especially given that COVID-19 has not yet been eradicated. The increase in the number of monkeypox cases bore some resemblance to the early stages of the COVID-19 pandemic. However, the timely dissemination of information through all media, without neglecting social networks, was crucial. Twitter played an important role as an open platform, allowing the discussion of issues related to public health. Following the performance of the sentiment analysis, as shown in Table 2 and Figure 8, the results indicate that the opinions of Twitter users can be mostly classified as neutral. It can be presumed that users do not have negative or fearful feelings, but that they simply tweeted, pushed by the conjunctural moment of the monkeypox virus. Very similarly, ref. [29] obtained very similar results regarding monkeypox. For example, they found that 60% of the posted tweets related to monkeypox lacked sentiment or corresponded to misinformation, whereas only 32% corresponded to reliable or useful information for processing and classification. The proposed CNN-LSTM model obtained modest results on accuracy, specificity, recall, and F1 score metrics, with averages of 83%, 99%, 85%, and 83%, respectively. Compared to the deep-learning-based approach of [27], better results were obtained in the same metrics of precision, recall, F1 score, and accuracy, with averages of 85.44%, 85.47%, 85.40%, and 87.13%, respectively. This is an indicator of the need to evaluate when choosing pre-trained models for classification and sentiment analysis. ML classification models generally perform better at higher data volumes, which helps to significantly improve accuracy. In comparison with the results obtained in this work, ref. [19] obtained very good results in the accuracy metric (96%), with an open-source automatic learning model based on the BERT algorithm.

Limitations

The limitations of this study include the rapid evolution of the monkeypox virus and the fact that the analysis and classification were based on Twitter posts by users from Peru, meaning that only tweets in Spanish were considered. Therefore, there is a lack of representativeness relative to other countries, and the results may change over time depending on how the monkeypox virus behaves. Additionally, these results do not necessarily represent the entire population.

5. Conclusions

In the field of sentiment analysis, there are various approaches and machine learning models pre-trained to perform these actions. However, it is useful to know which approach and model have the best rates in the measurement metrics. The training, regardless of the model, will depend on the size of the data set; the larger the data volume, the better the model is optimized. In this work, the hybrid CNN-LSTM model was used to perform sentiment analysis and determine the emotional polarity of the content in the tweets. The results obtained in this work are based on user perception regarding the monkeypox virus as positive, negative, or neutral. The model applied in this work, supported by the optimizers, demonstrated high average performance, achieving an average accuracy rate of 83%. According to the other metrics obtained, this is considered a relatively successful approach and suitable for classifying sentiments in tweets related to the monkeypox virus.

In Figure 8, we can see the classification of feelings, based on the data set of 84,018 tweets related to the monkeypox virus. These feelings were classified into the following categories: positive, negative, and neutral, representing 35.13%, 19.45%, and 45.42%, respectively. The results indicate that 45.42% of the users are not afraid of the monkeypox infectious disease.

However, 19.45% of the users expressed negative feelings, such as fear, risk, and death, and sought to avoid catching or acquiring the disease by any means of contagion.

The present work has theoretical and practical implications that can be very helpful for future work. Theoretically, we applied a machine learning approach with the hybrid CNN-LSTM model, integrating convolutional networks and a type of recurrent neural network, to optimize prediction results. Practically, this work can be useful for the community as a whole since it helps to raise awareness of the monkeypox virus. Finally, to improve model performance, higher data volume and more robust techniques should be used that can be refined with future research to improve accuracy rates.

Author Contributions: Conceptualization, M.C.-C.; methodology, O.I.-V.; validation, J.L.H.S.; formal analysis, S.B.-C. and J.A.Y.; investigation, J.Z.-P.; data curation, M.C.-C., A.A.-R. and J.A.Y.; writing—original draft preparation, M.C.-C., A.A.-R. and J.A.Y.; writing—review and editing, J.Z.-P., A.A.-R. and J.A.Y.; visualization, S.B.-C., A.A.-R. and J.A.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Parker, S.; Buller, R.M. A review of experimental and natural infections of animals with monkeypox virus between 1958 and 2012. *Future Virol.* **2013**, *8*, 129–157. [CrossRef] [PubMed]
2. Ladnyj, D.; Ziegler, P.; Kima, E. A human infection caused by monkeypox virus in Basankusu Territory, Democratic Republic of the Congo. *Bull. World Health Org.* **1972**, *46*, 593. [PubMed]
3. Organización Mundial de la Salud. Viruela Símica. 2022. Available online: <https://www.who.int/es/news-room/fact-sheets/detail/monkeypox> (accessed on 7 September 2022).
4. Mohbey, K.K.; Meena, G.; Kumar, S.; Lokesh, K. A CNN-LSTM-based hybrid deep learning approach to detect sentiment polarities on Monkeypox tweets. *arXiv* **2022**, arXiv:2208.12019. [CrossRef]
5. Thornhill, J.P.; Barkati, S.; Walmsley, S.; Rockstroh, J.; Antinori, A.; Harrison, L.B.; Palich, R.; Nori, A.; Reeves, I.; Habibi, M.S.; et al. Monkeypox Virus Infection in Humans across 16 Countries—April–June 2022. *N. Engl. J. Med.* **2022**, *387*, 679–691. [CrossRef] [PubMed]
6. Livemint. World Health Network Declares Monkeypox a Pandemic. Available online: <https://www.livemint.com/news/world/world-health-network-declares-monkeypox-a-pandemic-11655951549860.html> (accessed on 9 September 2022).
7. Hammarlund, E.; Lewis, M.W.; Hanifin, J.M.; Mori, M.; Koudelka, C.W.; Slifka, M.K. Antiviral Immunity following Smallpox Virus Infection: A Case-Control Study. *J. Virol.* **2010**, *84*, 12754–12760. [CrossRef]
8. Lytras, M.D.; Visvizi, A. Big Data Research for Social Science and Social Impact. *Sustainability* **2019**, *12*, 180. [CrossRef]
9. McCormick, T.H.; Lee, H.; Cesare, N.; Shojaie, A.; Spiro, E.S. Using Twitter for Demographic and Social Science Research: Tools for Data Collection and Processing. *Sociol. Methods Res.* **2017**, *46*, 390–421. [CrossRef]
10. Lokesh, S.; Kumar, P.M.; Devi, M.R.; Parthasarathy, P.; Gokulnath, C. An Automatic Tamil Speech Recognition system by using Bidirectional Recurrent Neural Network with Self-Organizing Map. *Neural Comput. Appl.* **2019**, *31*, 1521–1531. [CrossRef]
11. Kydros, D.; Argyropoulou, M.; Vrana, V. A Content and Sentiment Analysis of Greek Tweets during the Pandemic. *Sustainability* **2021**, *13*, 6150. [CrossRef]
12. Iparraguirre-Villanueva, O.; Guevara-Ponce, V.; Sierra-Linan, F.; Beltozar-Clemente, S.; Cabanillas-Carbonell, M. Sentiment Analysis of Tweets using Unsupervised Learning Techniques and the K-Means Algorithm. *Int. J. Adv. Comput. Sci. Appl.* **2022**, *13*, 571–578. [CrossRef]
13. Lorenzoni, V.; Andreozzi, G.; Bazzani, A.; Casigliani, V.; Pirri, S.; Tavošchi, L.; Turchetti, G. How Italy Tweeted about COVID-19: Detecting Reactions to the Pandemic from Social Media. *Int. J. Environ. Res. Public Health* **2022**, *19*, 7785. [CrossRef] [PubMed]
14. Hakak, S.; Alazab, M.; Khan, S.; Gadekallu, T.R.; Maddikunta, P.K.R.; Khan, W.Z. An ensemble machine learning approach through effective feature extraction to classify fake news. *Future Gener. Comput. Syst.* **2020**, *117*, 47–58. [CrossRef]
15. Iparraguirre-Villanueva, O.; Guevara-Ponce, V.; Paredes, O.R.; Sierra-Liñan, F.; Zapata-Paulini, J.; Cabanillas-Carbonell, M. Convolutional Neural Networks with Transfer Learning for Pneumonia Detection. *Int. J. Adv. Comput. Sci. Appl.* **2022**, *13*. [CrossRef]

16. Gal, Y.; Ghahramani, Z. A Theoretically Grounded Application of Dropout in Recurrent Neural Networks. *Adv. Neural Inf. Process Syst.* **2016**, *29*, 1027–1035. [[CrossRef](#)]
17. Zain, Z.M.; Alturki, N.M. COVID-19 Pandemic Forecasting Using CNN-LSTM: A Hybrid Approach. *J. Control. Sci. Eng.* **2021**, *2021*, 8785636. [[CrossRef](#)]
18. World Health Organization. Multi-Country Monkeypox Outbreak: Situation Update. 2022. Available online: <https://bit.ly/39hcWiQ> (accessed on 11 September 2022).
19. Kolluri, A.; Vinton, K.; Murthy, D. PoxVerifi: An Information Verification System to Combat Monkeypox Misinformation. Available online: <http://arxiv.org/abs/2209.09300> (accessed on 11 September 2022).
20. Xie, Y.; Xing, F.; Kong, X.; Su, H.; Yang, L. Beyond classification: Structured regression for robust cell detection using convolutional neural network. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; Volume 9351, pp. 358–365.
21. Yamashita, R.; Nishio, M.; Do, R.K.G.; Togashi, K. Convolutional neural networks: An overview and application in radiology. *Insights Imaging* **2018**, *9*, 611–629. [[CrossRef](#)] [[PubMed](#)]
22. Ayyildiz, E.; Erdogan, M.; Taskin, A. Forecasting COVID-19 recovered cases with Artificial Neural Networks to enable designing an effective blood supply chain. *Comput. Biol. Med.* **2021**, *139*, 105029. [[CrossRef](#)]
23. Daradkeh, M. Analyzing Sentiments and Diffusion Characteristics of COVID-19 Vaccine Misinformation Topics in Social Media: A data analytics framework. *Int. J. Bus. Anal.* **2022**, *9*, 22. [[CrossRef](#)]
24. Vadivel, S.; Jayakarthy, R. Predictive analytics on Covid using recurrent neural network. In *AIP Conference Proceedings*; AIP Publishing LLC: New York, NY, USA, 2022; Volume 2393, p. 020194. [[CrossRef](#)]
25. Ma, J.; Wang, L.; Zhang, Y.-R.; Yuan, W.; Guo, W. An integrated latent Dirichlet allocation and Word2vec method for generating the topic evolution of mental models from global to local. *Expert Syst. Appl.* **2023**, *212*, 118695. [[CrossRef](#)]
26. Ortiz-Martínez, Y.; Sarmiento, J.; Bonilla-Aldana, D.K.; Rodríguez-Morales, A.J. Monkeypox goes viral: Measuring the misinformation outbreak on Twitter. *J. Infect. Dev. Ctries.* **2022**, *16*, 1218–1220. [[CrossRef](#)]
27. Gandomi, A.; Haider, M. Beyond the hype: Big data concepts, methods, and analytics. *Int. J. Inf. Manag.* **2015**, *35*, 137–144. [[CrossRef](#)]
28. Sitaula, C.; Shahi, T.B. Monkeypox virus detection using pre-trained deep learning-based approaches. *J. Med. Syst.* **2022**, *46*, 78. [[CrossRef](#)]
29. Li, M.; Ch'Ng, E.; Chong, A.Y.L.; See, S. Multi-class Twitter sentiment classification with emojis. *Ind. Manag. Data Syst.* **2018**, *118*, 1804–1820. [[CrossRef](#)]
30. Rahmanian, V.; Jahanbin, K.; Jokar, M. Using twitter and web news mining to predict the monkeypox outbreak. *Asian Pac. J. Trop. Med.* **2022**, *15*, 236. [[CrossRef](#)]

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.