



## OPEN ACCESS

## EDITED BY

Zisis Kozlakidis,  
International Agency for Research on Cancer  
(IARC), France

## REVIEWED BY

Belfin R. V.,  
Karunya Institute of Technology and  
Sciences, India  
Marie B. Romond,  
Université Lille 2 Droit et Santé, France

## \*CORRESPONDENCE

Eugenia Bezirtzoglou  
✉ empezirt@yahoo.gr

†These authors have contributed equally to this work and share first authorship

RECEIVED 22 March 2023

ACCEPTED 30 June 2023

PUBLISHED 18 July 2023

## CITATION

Stefanis C, Giorgi E, Kalentzis K, Tselemonis A, Nena E, Tsigalou C, Kontogiorgis C, Kourkoutas Y, Chatzak E, Dokas I, Constantinidis T and Bezirtzoglou E (2023) Sentiment analysis of epidemiological surveillance reports on COVID-19 in Greece using machine learning models. *Front. Public Health* 11:1191730. doi: 10.3389/fpubh.2023.1191730

## COPYRIGHT

© 2023 Stefanis, Giorgi, Kalentzis, Tselemonis, Nena, Tsigalou, Kontogiorgis, Kourkoutas, Chatzak, Dokas, Constantinidis and Bezirtzoglou. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Sentiment analysis of epidemiological surveillance reports on COVID-19 in Greece using machine learning models

Christos Stefanis<sup>1†</sup>, Elpida Giorgi<sup>1†</sup>, Konstantinos Kalentzis<sup>1</sup>, Athanasios Tselemonis<sup>1</sup>, Evangelia Nena<sup>2</sup>, Christina Tsigalou<sup>3</sup>, Christos Kontogiorgis<sup>1</sup>, Yiannis Kourkoutas<sup>4</sup>, Ekaterini Chatzak<sup>5</sup>, Ioannis Dokas<sup>6</sup>, Theodoros Constantinidis<sup>1</sup> and Eugenia Bezirtzoglou<sup>1\*</sup>

<sup>1</sup>Laboratory of Hygiene and Environmental Protection, Department of Medicine, Democritus University of Thrace, Alexandroupolis, Greece, <sup>2</sup>Pre-Clinical Education, Laboratory of Social Medicine, Medical School, Democritus University of Thrace, Alexandroupolis, Greece, <sup>3</sup>Laboratory of Microbiology, Medical School, Democritus University of Thrace, Alexandroupolis, Greece, <sup>4</sup>Laboratory of Applied Microbiology, Department of Molecular Biology and Genetics, Democritus University of Thrace, Alexandroupolis, Greece, <sup>5</sup>Laboratory of Pharmacology, Medical School, Democritus University of Thrace, Alexandroupolis, Greece, <sup>6</sup>Department of Civil Engineering, Democritus University of Thrace, Komotini, Greece

The present research deals with sentiment analysis performed with Microsoft Azure Machine Learning Studio to classify Facebook posts on the Greek National Public Health Organization (EODY) from November 2021 to January 2022 during the pandemic. Positive, negative and neutral sentiments were included after processing 300 reviews. This approach involved analyzing the words appearing in the comments and exploring the sentiments related to daily surveillance reports of COVID-19 published on the EODY Facebook page. Moreover, machine learning algorithms were implemented to predict the classification of sentiments. This research assesses the efficiency of a few popular machine learning models, which is one of the initial efforts in Greece in this domain. People have negative sentiments toward COVID surveillance reports. Words with the highest frequency of occurrence include government, vaccinated people, unvaccinated, telephone communication, health measures, virus, COVID-19 rapid/molecular tests, and of course, COVID-19. The experimental results disclose additionally that two classifiers, namely two class Neural Network and two class Bayes Point Machine, achieved high sentiment analysis accuracy and F1 score, particularly 87% and over 35%. A significant limitation of this study may be the need for more comparison with other research attempts that identified the sentiments of the EODY surveillance reports of COVID in Greece. Machine learning models can provide critical information combating public health hazards and enrich communication strategies and proactive actions in public health issues and opinion management during the COVID-19 pandemic.

## KEYWORDS

sentiment, classification, COVID-19, Facebook, public health, machine learning, natural language processing

# 1. Introduction

The pandemic crisis that broke out due to the COVID-19 disease not only tested the limits of the health systems in all countries but also other aspects of the citizens' political, economic and social life. Dissemination of information about the hazard of COVID-19, infections, deaths and the respective measures and actions toward mitigating consequences by each country has been considered of great importance (1, 2).

The enormous amount of data produced in the era of COVID has manifested a new research path that is; the intensification of Big Data analytics and Artificial Intelligence toolbox, mainly the implementation of machine learning algorithms and predictive modes. The goal is to apply models in order to predict the risk, to forecast COVID pandemic waves, to diagnose results, and to detect information and misinformation activities related to public health management in Social Networking Sites (SNS), as well the society (3).

One of the main communication channels of public authorities during a crisis, like a health crisis, is social media. Especially since, in today's era, the internet offers speed and immediacy in disseminating critical information from the state to the citizens. Planning and drawing up a communication strategy through social media is critical. It requires a high degree of expertise to not produce opposite results, such as misinformation and confusion among citizens. Therefore, social media became a powerful weapon in state and health agencies' arsenal against the pandemic crisis, intending to inform citizens at individual and state levels about the evolution of the pandemic, individual protection measures, restrictive measures and upcoming health policies which were to be followed (4, 5).

In addition, issues and policies related to the evaluation of health services, the dissemination of preventive health, psychological and educational actions against COVID-19 were disseminated to the public and extensively commented on through Social Networking Sites. The next step in utilizing these online SNS is knowledge as well as data mining for early warning and detection of COVID-19 incidents, the emotional orientation of the public regarding the whole range of actions during the pandemic crisis, and filtering of information and misinformation from the public (6, 7).

Sentiment analysis can decipher people's opinions, emotions and sentiments. Political sciences, marketing and sales use sentiment analysis to improve products and services following up customer reviews. In the medical field, sentiment analysis is used to spot and extract opinions and sentiments on social media on issues regarding mental health, epidemiology, new medical treatments, drugs and supplements, patient forums, and pharmacovigilance (8).

The use of social media should also be discussed since sentiment orientation analysis utilizes such data originating from social discussions. However, why is social media used in public health matters? The answer is conclusively given by the fact that critical information can be extracted from these, highlighting the population's demographic, spatial and socio-economic disparities. However, using these data and big data for infectious disease surveillance should be evaluated for their reliability and credibility in research initiatives concerning public health policy, public

opinion and trends, health crisis management, pandemics and infectious disease control and surveillance (9, 10).

Health issues surveillance is an additional role of SNS in their usefulness as a knowledge tool in public health. Moreover, analyzing features and generating real time data of pharmacovigilance, information or misinformation surveillance, and tracking health behavior issues like parents' health literacy skills are also embodied in sentiment analysis in the public health domain (11–13).

In the middle of the last decade, the Centers for Disease Control and Prevention (CDC) has recognized the role of health-related data that signal a possibility of an outbreak or the initial development of an epidemic outbreak and can signify a disturbance in public health (syndromic surveillance<sup>1</sup>). The speed of exploiting this data with the right tools, such as sentiment analysis, due to a 1–2 weeks lag between diagnosis and when this information becomes part of published statistics, has resulted in a practical approach to epidemic detection. Subsequently, the formation or combination of surveillance systems aimed at detecting a particular disease and prompting health officials to take appropriate measures and communicate health policies to prevent outbreaks can be realized (14).

National Public Health Organization (EODY) in Greece supervise all the public health services related to communicable and chronic diseases and implement all the necessary actions for epidemiological surveillance, risk assessment, scientific consultation, dissemination and communication strategies to inform the population about health issues, like COVID-19 surveillance reports and data (15). The main innovations of our research include the following:

- Analyze the sentiment orientation of Greek citizens' comments regarding the surveillance reports of COVID-19 during the last winter of the pandemic (December 2021-January 2022).
- Determine various issues and topics the public discussed on EODY Facebook page.
- Apply machine learning algorithms to predict the sentiment orientation of public opinion toward an even more concise picture of the given period of time.
- Study people's emotions in Facebook posts to recognize and interpret how classification models can support the understanding of public perception toward the COVID-19 pandemic.

The current contribution of this research approach is to develop and propose a basis for identifying issues about the impact of the surveillance reports of COVID. Extracting public sentiments from Facebook provides the raw material to decision makers in Greek public health agencies for compassing new communication strategies based on digital communications via social media and public reaction monitoring.

1 <https://search.cdc.gov/search/?query=syndromic%20surveillance&page=1>

## 1.1. Related work

In our effort to identify and highlight the applications of sentiment analysis in the field of public health and the management of infectious diseases, pandemics and the combination of such tools in existing public health management and surveillance systems, we conducted a search on Scopus for research papers. The aim is 2-fold: to identify works in the above field, namely the combination of sentiment analysis with related initiatives in public health. Additionally, we realize science mapping to reveal the research content of such approaches by running a co-occurrence analysis, a bibliometric tool and spotlighting the linkage of sentiment analysis and its use in public health (Supplementary Figure 1).

The phrase “infectious AND disease AND sentiment” in the TITLE-ABS-KEY search field was used for scanning the Scopus database and the bibliometric approach. The year 2023 was excluded, while only English manuscripts and various document types were included. The VOS Viewer software was applied to visualize the results and create a bibliographic map<sup>2</sup> (Figure 1). The full counting method was further used in the co-occurrence analysis of the keywords in the title, abstract, and text of the manuscripts. Briefly, 1,380 keywords were initially extracted, but setting the minimum number of occurrences to 5, 66 were included. The bibliographic network revealed 4 clusters with 54 keywords on the map (Supplementary Figure 1) (16, 17). Keywords, clusters, weight score and number of occurrences of each item in the bibliographic map are listed in Supplementary Table 2.

As stated above, the bibliographic map categorizes terms into four significant clusters with respective colors. The color of each word is dictated by the cluster to which it belongs. Moreover, the closer two terms are spotted on the map, the stronger their relatedness. The yellow cluster contains words referring to the COVID-19 pandemic crisis, such as COVID-19, pandemic, coronavirus and coronavirus infection. The green cluster includes terms related to vaccines and vaccination. The blue cluster circulates items like infectious disease, disease surveillance, communicable disease and disease transmission. At the same time, the red one encompasses terms like sentiment analysis, infectious disease, data mining, and social media platforms.

Interpreting the bibliographic network and the included keywords, one should note the engagement and the boost that the COVID period offered in the surveillance, management and monitoring of infectious diseases via social media and online media platforms with computational linguistic applications like sentiment analysis. Such applications were already in use; however, the pandemic period spotlighted extra interest and research value. Public perception, opinions, patients' experience, opinion mining and sentiment analysis are established as vital tools in infoepidemics, infodemiology, and infoveillance's arsenal (18–22).

In the current literature regarding sentiment analysis studies concerning public health, COVID-19 and data generated and circulated in social media focused on data extracted from Twitter. A proposed methodological framework addressed the problem of public fear during the COVID-19 peak in the United States of America. The characteristics of this framework are based on

machine learning algorithms, namely logistic regression and Naïve Bayes, as well as text analysis methods (23).

In the past several years, COVID-19 has revolutionized many aspects of everyday life. One aspect is the working environment, especially in developed countries, where remote work has become common practice. This transition to new labor standards at the commencement of the world epidemic using sentiment analysis was the goal of another contribution (24).

Public sentiments followed the deviation of the three pandemic waves in Croatia, a fact revealed in a recent study. This study underlined the significant impact of COVID-19 on the psychological side effects of society. Twitter data, sentiment analysis, and machine learning algorithms were implemented to detect, among others, the polarity of Croatian public opinion (25).

A research study aimed at analyzing Twitter data on citizens' attitudes toward vaccination policies and issues in the USA. The findings revealed positive attitudes toward vaccination and necessary safety measures against COVID-19 (26). Furthermore, the justification of sentiment analysis using data from SNS is also manifested in another proposed research. The outcome emphasized that emotional state and sentiment polarity information during the pandemic can support communication strategies and public guidance (27).

Further research focused on Twitter data, COVID-19 pandemic and governmental management actions by studying public sentiment orientation and emotional positions (28). Besides the sentiment analysis, this research team developed a decision support system to scrutinize Facebook posts to aid the decision making of public health agencies during the diverse health crisis of COVID-19.

In recent years, the sentiment analysis emerging from text mining of various Social Networking Sites has been increasing in Greek scientific literature in multiple research fields like informatics and management, business administration, political sciences, computer engineering and statistics (29–35). In the field of public health, there are very few research efforts and even fewer studies focusing on COVID-19 and Facebook comments (7, 36). Considering the data sources, social media platforms, disease-specific communities, patient forums, blogs, and electronic health records and platforms are the most common. Among the most common platforms for the exploitation of health and patient experience data is Twitter (37, 38). The primary usage of the Twitter platform in health-related issues like public health and infectious diseases is for content analysis and secondary for surveillance (39). The prospective and promising usage of the Internet data sources in infectious disease and epidemic surveillance is realized increasingly as an effective tool. Data's spatial and temporal distribution can be utilized for internet-based surveillance, disease forecasting and modeling (40).

In the healthcare industry, opinion mining and sentiment extraction are vital. Patients express their opinions and sentiments regarding old and contemporary treatments, medicines and public health services. In this vein, a research study highlights the analysis of tweets concerning diabetes. Additionally, a health web forum was utilized to extract the sentiments of people and the differentiation of genders diagnosed with HIV and the respective issues confronting their lives (41, 42).

2 <https://www.vosviewer.com/>

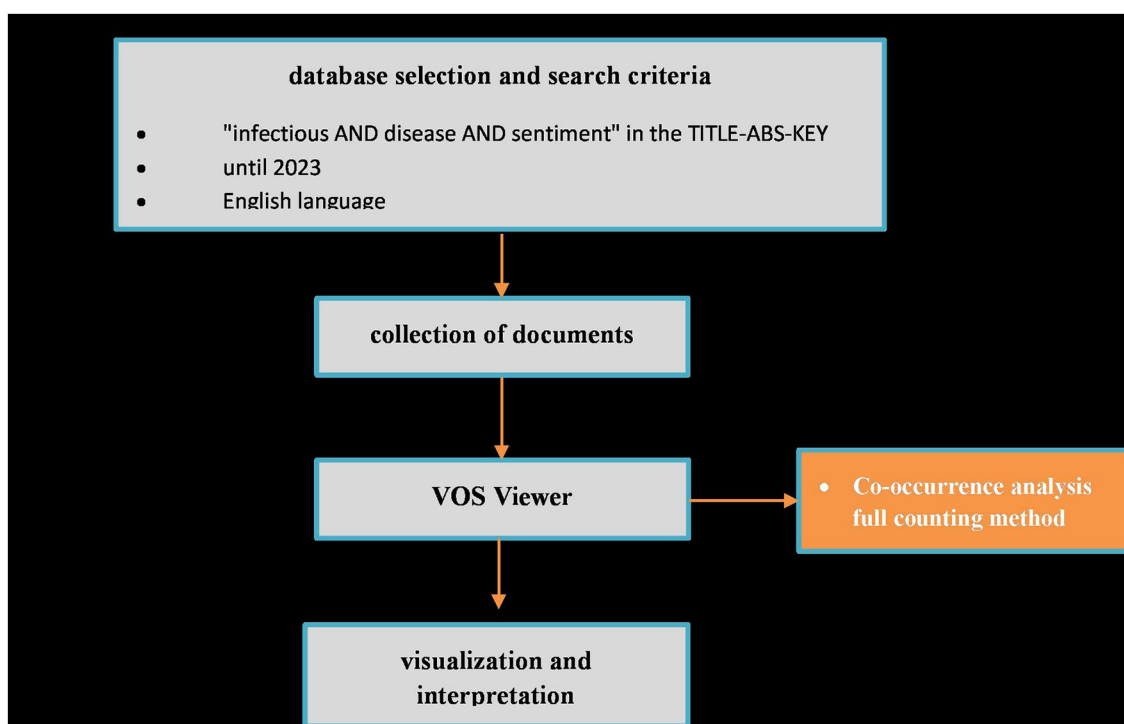


FIGURE 1  
Research workflow diagram of bibliometric analysis.

## 2. Materials and methods

### 2.1. Overall methodological approach

The analysis of sentiment orientation includes a sequence of steps to ensure the correct depiction of the properties evaluated based on the categorization of emotions. Briefly, **collecting data** from the source, which in this particular study is **Facebook**, is the first step. Next is data **preprocessing**, which usually refers to removing words, grammar, and syntactical phrases to clean up a sentence, and finally the attribution of the sentiment based on the method followed, the visualization and the interpretation of the results. Regarding the sentiment analysis technique, there are two axes: the process based on machine learning and the one established in a lexicon. The hybrid method combines the primary two techniques (43).

The core of sentiment determination is distinguished in a binomial characterization, positive or negative. There can also be a third category, i.e., positive, negative or neutral emotional content. Essentially, with the machine learning methods, the problem of the sentiment attribution of comments has resulted in a data categorization problem. By extension, classification algorithms are usually applied, such as, among others, Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), Stochastic Gradient Descent (SGD), and ensemble classifiers.

Logistic regression can be applied to classification tests by predicting the binary dependent variable from a set of independent variables. Another binary, non-probabilistic classifier is the support vector machine which relies on kernel mapping. Moreover, the

random forest algorithm produces multiple trees where each one of them is constructed using a random subset of the vector features. The decisions of each tree are synthesized utilizing an algorithm that gives the outcome. Additionally, two class machine learning classifiers implemented in this study which are: the two class neural network, the two class Bayes point machine, the two class boosted decision tree, the two class averaged perceptron, the two class decision jungle, the two class local deep SVM, and all these algorithms develop a binary classification model (44, 45).

The overall statistical conduct of machine learning classifiers is appraised with the aid of respective parameters, the most popular being the accuracy (1), precision (2), recall (3), and F1 score (4). These metrics are composed of TP, TN, FP, and FN values representing true positive, true negative, false positive, and false negative values in a produced confusion matrix (42, 44). The equations of the evaluation parameters are:

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

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

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

$$F1\ score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

Natural language processing (NLP) is a subfield of computational science and refers to a software's capability to automatically manipulate and classify information using natural language, like text and speech. Sentiment analysis was based on a Natural



Language Processing algorithm, a subset of Artificial Intelligence (AI). Contemporary Natural Language Processing methods, like sentiment analysis classifiers and models, have been successfully utilized to fight the COVID pandemic (45).

## 2.2. Methodology applied in the case study

A machine learning technique was performed with the Azure Excel Add-in to classify Facebook data (comments) from the official page of EODY. In particular, daily reports publicized from November 2021 to January 2022 were collected. Moreover, after processing 300 comments, positive, negative and neutral sentiments were included. This approach involved analyzing the comments of the people discussing the daily EODY reports on Facebook, which included, among others, the number of deaths, number of patients in the critical condition/intensive care, patient gender, and number of COVID-19 confirmed cases based on rapid and molecular tests (46, 47).

The Chi square independence test was used to evaluate sentiment orientation across gender and month. That is, to examine whether Facebook comment frequencies significantly differed across men and women and across the three months of the last winter of the pandemic in Greece, namely, November 2021-January 2022. The null hypothesis in each case is that the two genders do not comment independently of one another. Furthermore, the sentiment orientation of the comments varies somewhat regarding the month they appeared. Results with a  $p$ -value  $<0.05$  were considered statistically significant.

In the next phase, all neutral comments were omitted, and the remained comments were corrected manually to leverage the machine learning performance, especially in the case of sarcasm and irony in Greek phrases. Finally, only the positive and negative comments were selected to create and compare machine learning classifiers, namely 199 comments. Subsequently, models were developed in Microsoft Machine Learning Studio (Classic) to classify public sentiments, positive or negative, based on each surveillance report on EODY's official Facebook page (Figure 2). Statistical analysis was performed by SPSS v.27 statistical software (SPSS Inc., USA) and Microsoft Excel (48).

This research tests nine classification techniques and the respective evaluation metrics, namely two-class classifiers: Neural Network, Bayes Point Machine, Decision Forest, Boosted Decision Tree, Decision Jungle, Locally Deep Support Vector Machine, Logistic Regression, Support Vector Machine and Averaged Perceptron (Figure 2). Moreover, cross validation techniques strengthen the part of the development of mathematical models and avoid overfitting or underfitting problems. This technique is followed in machine learning to evaluate the reliability of a model trained from a dataset and control data variability (49).

Specifically, the K-fold cross validation technique was adopted because it is one of the most common approaches (50–54). The model is trained using an exclusive combination of K-1 subsets of data and tested on the remaining subset. Briefly, the training dataset is divided into K subsets of equal size, which in this study equals ten. Subsequently, ten models will be generated for each subset of training data and evaluated by averaging the performance metric

values of the models, i.e., accuracy, precision, recall and F1 score<sup>3</sup> (Figure 2).

## 3. Results and discussion

The text analysis revealed positive, negative and neutral sentiments expressed on EODY's Facebook page in 3 months, from November 2021 to January 2022 (Figures 3A, B). The vast majority of the sentiments have a negative orientation, with the respective percentage rising to 57%. On the contrary, 34% neutral and 9% positive comments were also documented (Figure 3B).

A declined trend of negative, neutral and positive sentiments at the end of the winter, in January 2022, is also highlighted in Figure 3A. The interpretation of this downward trend of all types of comments stems from the revision of EODY's dissemination policy. More concretely, the organization decided to publish weekly epidemiological reports of COVID-19 during the first ten days of January instead of the daily publication of the records. Consequently, the public lost interest, and citizens' comments on the EODY Facebook page decreased.

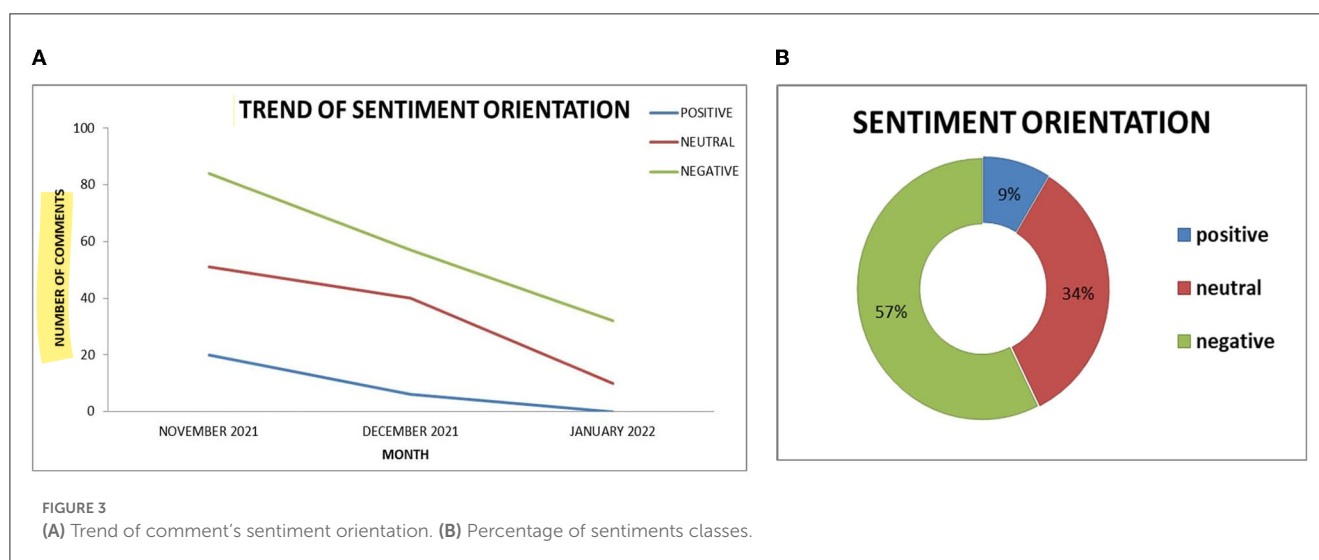
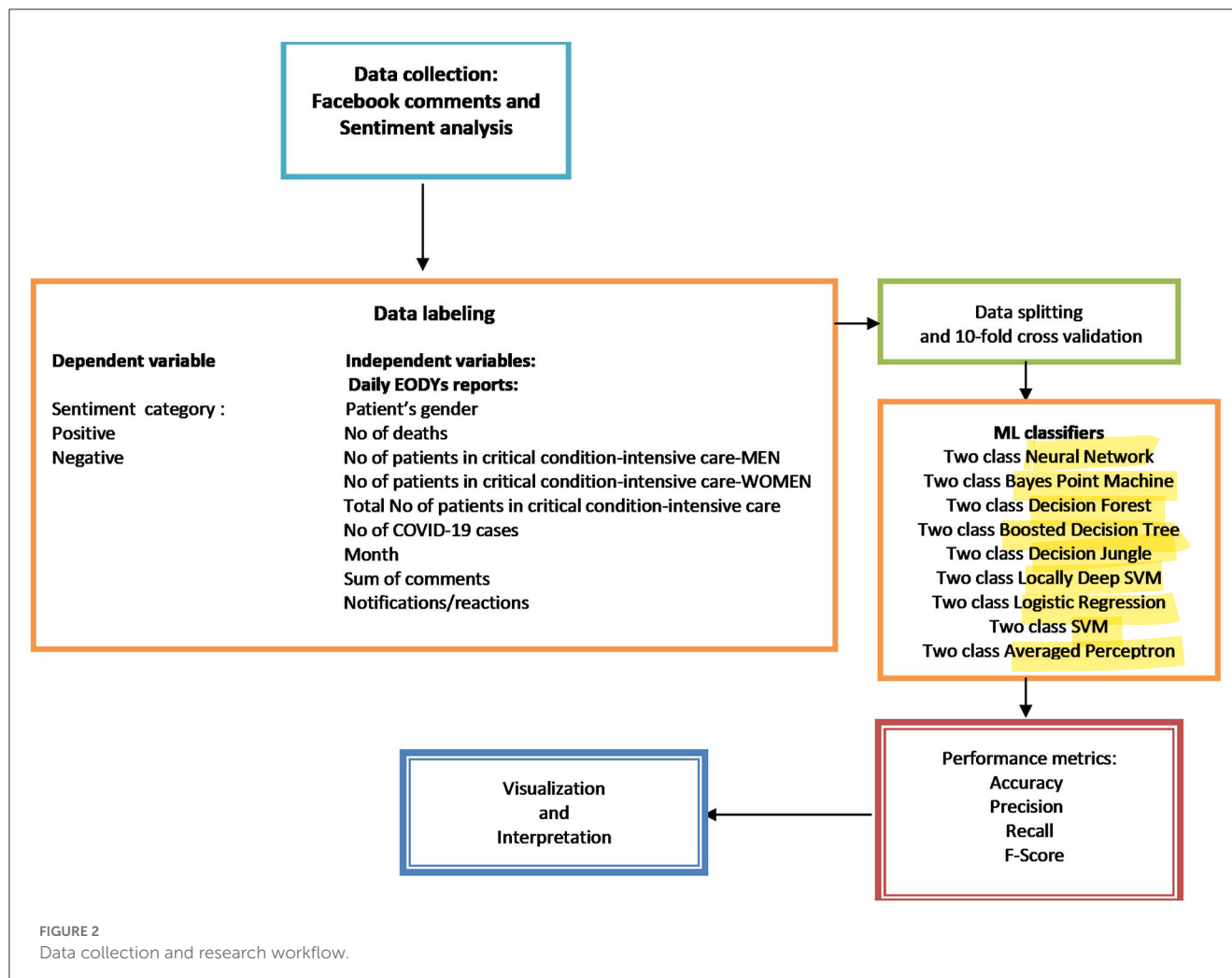
The variation in sentiment orientation depends on the course of the pandemic and the corresponding pandemic waves that contribute to the public's positive or negative comments (25). A related study reported sentiment analyses of COVID-19 Facebook posts from January 1, 2019, to March 18, 2020. More precisely, comments were collected from three Facebook public health agencies' pages, the Ministry of Health in Singapore, the Centers for Disease Control and Prevention in the United States, and Public Health England in England. A negative polarity was indicated for most of the comments in all three public agencies. Moreover, the temporal analysis showed variations between the number of posts in three countries, partly explained by the use of Facebook by the health services to publicize information and topics on COVID-19 and other health issues (55).

Similarly, the high negative percentage of public opinion during the pandemic was in line with other research outcomes; namely: averaged positive, negative, and neutral sentiments were at 58%, 22%, and 17% in the United Kingdom, 56%, 24%, and 18% in the United States, respectively (56).

Table 1 highlights the sentiment proportion of each comment for every month of the examination period, namely November 2021, December 2021 and January 2022. Approximately half of the Facebook posts (51.7%) were formulated in November, while the lowest percentage of comments was documented in January. Overall, January noted the lowest percentage of Facebook posts in all three categories, mainly positive (0%), negative comments (10.7%), and neutral (3.3%).

Furthermore, it is worth noting that the continuous decrease in the number of posts throughout the 3 months was statistically significant ( $p < 0.05$ ). As mentioned previously, this decline can be attributed to the decision made by EODY (National Public Health Organization in Greece) to change the frequency of publishing COVID-19 surveillance reports from daily to weekly. This alteration in the reporting schedule directly impacted the

3 <https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/cross-validate-mode>



quantitative production of Facebook posts among the public, reducing overall activity on the platform over time.

During the initial phase of the pandemic crisis in China, specifically between mid-January and mid-February, there was

a noticeable rise in negative comments on a social networking platform. This increase could be attributed to the official confirmation and documentation of human-to-human virus transmission during that period. The revelation of this critical

TABLE 1 Percentage of sentiments per month.

Sentiment* Month <sup>1</sup>	November	December	January
Negative	28.0%	19%	10.7%
Neutral	17%	13.3%	3.3%
Positive	6.7%	2%	0.0%
Total	51.7%	34.3%	14%

<sup>1</sup> $\chi^2 = 12,786$ ,  $df = 4$ ,  $p < 0.05$ . \*Correlation.

TABLE 2 Percentage of sentiments per gender.

Sentiment* Gender <sup>1</sup>	Men	Women
Negative	31.3%	26.3%
Neutral	17.0%	16.7%
Positive	6.3%	2.3%
Total	54.7%	45.3%

<sup>1</sup> $\chi^2 = 4,273$ ,  $df = 2$ ,  $p > 0.05$ . \*Correlation.

information likely impacted public sentiment, leading to a surge in adverse reactions and discussions on the platform. A survey conducted on the time trend of 500 tweets based on Vader Lexicon in August 2021 found that neutral comments accounted for 20% of the total, followed by negative comments at 17% and positive comments at 15%. However, in September of the same year, there was a slight increase in positive comments by 2%. On the other hand, the percentages of neutral and negative comments remained relatively stable, with no significant changes compared to the previous month (27, 57).

The sentiment patterns varied between genders. Specifically, men (31.3%) tended to make negative comments more than women (26.3%). Moreover, men expressed positive comments nearly three times more frequently than women. In summary, the data indicates that men, in general, contribute a higher number of comments compared to women. Moreover, men tend to produce a greater proportion of negative comments compared to women. However, when it comes to neutral comments, the percentages are relatively similar, with men having a slight predominance.

Women demonstrated a higher frequency of engaging with the healthcare system and had different evaluations of the health system regarding the COVID-19 crisis during the examination period in Greece (as shown in Table 2). Interestingly, the sentiment orientation of comments on EODY's Facebook page was not significantly influenced by gender ( $p > 0.05$ ).

Similar findings were observed when examining gender differences in expressing positive and negative sentiments on a health web forum. For instance, females used positive words such as "thank" and "glad" twice as often as men, while negative words like "problem," "scary," and "illness" were also used twice as frequently by females. On the other hand, males used positive words such as "important" and "receptive" more regularly, and negative words like "issue," "fever," and "aches" were used twice as often by males (42, 47). According to the survey results, it was found that men sent a higher number of messages compared to women. Specifically, men accounted for ~40% of the total messages analyzed, whereas

women constituted around 27% of the messages. This discrepancy in message participation between genders was observed in a dataset of nearly 23,000 messages (42).

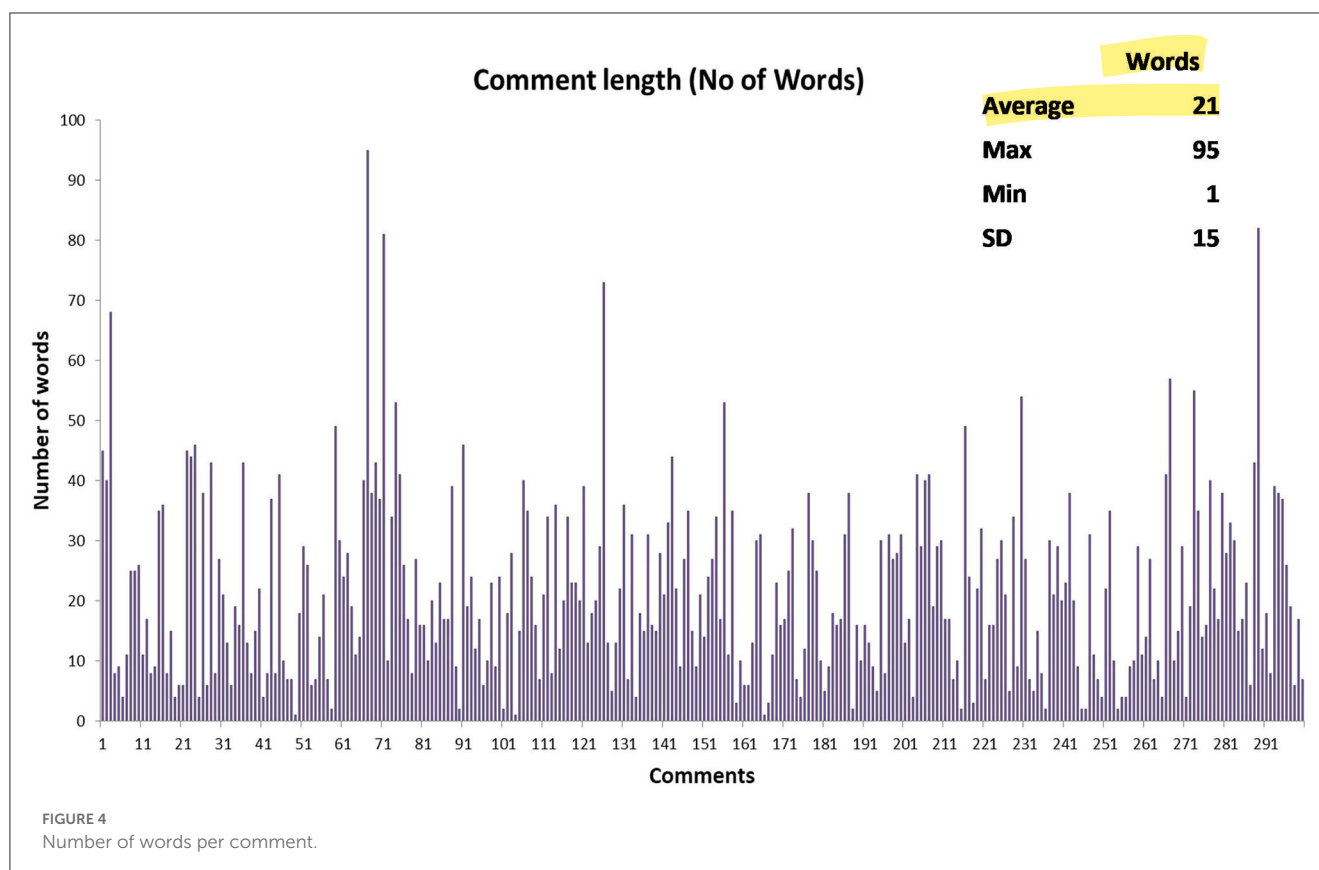
Figure 4 outlines the number of words composed for each comment. In total, 300 Facebook posts were collected from the EODY page. One comments consisted of one word, while one comment comprised of 95 words. The average post length was 21 words.

Figure 5 illustrates the words with a higher frequency rate in the public comments, reflecting opinions and debating issues, the greater the word the higher the frequency. Overall, 59 words are depicted. Words with the highest frequency of occurrence include "government," "vaccinated people," "unvaccinated," "telephone communication," "health measures," "virus," "COVID-19 rapid/molecular tests," "sad" and as expected, "COVID-19." Additionally, the quality of public data contained in the epidemiological reports published by EODY, as well as the responsibility toward the direction of reduction of coronavirus cases, are also the focus on citizens' debate.

The public discussed topics related to the lack of communication and guidance to people infected with COVID-19, especially on the first day after the positive result of a rapid or molecular test. During the pandemic, words like COVID, vaccines and all the related derivatives like corona, pandemic, infection, test, and measures are some of the most widely discussed. In the same context, the pandemic waves modified the terms that were commented among the public. People were interested in the virus early in the COVID-19 pandemic. Afterwards, public opinion focused on government measures, hygiene, and social and financial terms (27). The communication strategy of EODY and social media activity on Facebook are also criticized, while transparency issues emerged (58, 59).

Along the same line, public health and physical distancing are among the ten most discussed topics during WHO's press conferences during the pandemic. In addition, other hot topics like vaccine manufacture, contact tracing, report case, mild case, severe diseases, vaccination coverage, social measure, global solidarity, health emergency programme, disease control, and lock-down were underlined. Transparency issues also emerged during a survey regarding the UK government's COVID-19 control strategy in the first wave (April 2020). Citizens with various demographic and socioeconomic backgrounds were skeptical about the justification of opacity due to the pandemic, mistrust of politics, scientific evidence about the pandemic, the communication strategy and the decision-making processes implemented by the officials (60).

The word cloud analysis revealed that specific terms unrelated to sentiments appeared prominently, indicating the topics widely discussed among the public. One notable topic was vaccines and vaccinations, including discussions on public hesitancy, knowledge gaps, and misinformation. Another unique pattern was that younger individuals prioritized vaccination as a top concern compared to the older population, who relied on their previous vaccination experiences for other diseases. Transparency in government actions, vaccine manufacturing, and the lack of public trust, particularly among those associated with the anti-vaccination movement, were also significant issues. These findings emphasize the importance of the government's effective communication management and



strategic planning to address these concerns and build public trust (61–65).

In a separate survey conducted on the impact of COVID-19 in India, a country heavily affected by the pandemic, similar words such as “COVID” and “stay home” were observed. Additionally, topics related to hashtags such as #Lockdown, #COVID, and #corona were frequently discussed. The survey also revealed discussions on issues about working from home. These findings indicate that specific themes and concerns were shared across different studies conducted in different regions, highlighting the global impact and shared experiences related to the COVID-19 pandemic (66).

An analogous study aimed to analyze discussion topics from March 7 to April 21, 2020, and to perform sentiment analysis on 4 million Tweets. This probe revealed that people were interested in issues related to the confirmed cases and death rates, health authorities and government policies and adverse psychological reactions or psychological consequences (62). Negative emotions during the pandemic, such as anger, fear or sadness, were also justified in a study that determined the sentiment orientation at the global level from tweet posts from January 28 to April 9, 2020. The findings indicated that public health agencies should include measures toward leveraging citizens’ negative emotions and implement new actions in general hygiene communication management (67).

One aspect recognized during the COVID pandemic is mental health because religion and the spiritual factor are resilient factors. A research work underlined that factor with the aid

of interviews and natural language algorithms. The outcome spotlighted the positive effect of religion in human resilience during the symptoms of the disease. Moreover, words like security, confidence, tranquility, and peace were among the most stated between groups (68, 69).

In Figure 6, the comparison of the proposed classifiers is outlined. Two class Neural Network and Bayes Point Machine, were found to perform better than the other classifiers. Particularly, these two had the same score in Accuracy (87%), F1 score (36 and 35%), Precision (25 and 23%) and Recall (67 and 78%). The next classifier with the best score was the Logistic Regression classifier with 87%, 27%, 16%, and 78% metric evaluation values, respectively.

Other metrics should also be considered to estimate the predictive models’ performance. The precision value for a class is the number of true positives divided by the total number of items marked as belonging to the positive class. Recall in this context is defined as the number of true positives divided by the total number of items that belong to the positive category. In our analysis, where positive sentiment has been labeled as a “positive” and negative sentiment as “negative”, it is illustrated that the lowest Precision value is achieved by the Decision Forest and Boosted Decision tree, while the highest, 33%, by the Averaged Perceptron and the SVM. In other words, these last two algorithms have the best predictive value of positive emotions.

The highest Recall score was performed from the Logistic Regression and Bayes Point Machine classifiers, meaning that these two algorithms are more sensitive than the other classifiers. Finally,



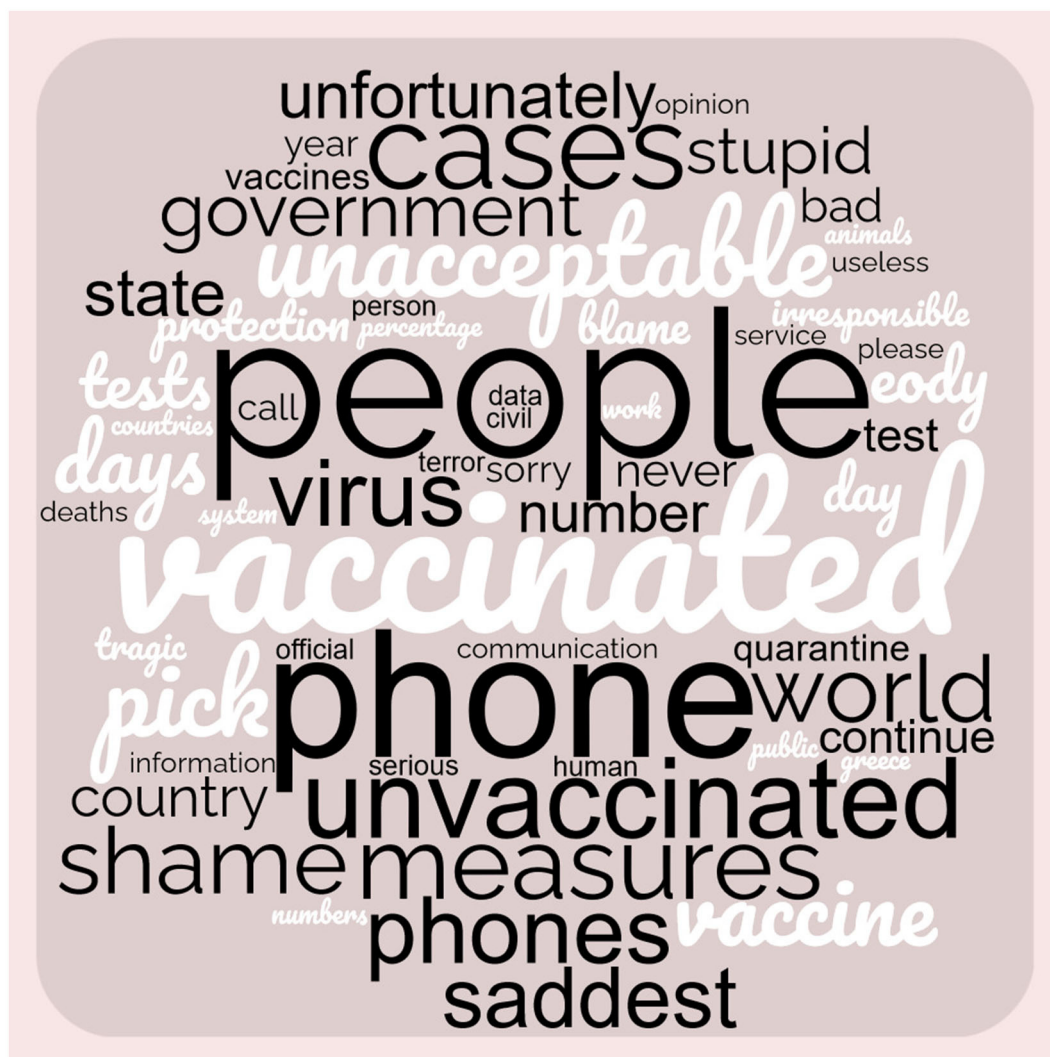


FIGURE 5

The word cloud represents the frequency of word occurrence and serves as a depiction and evaluation of the audience's perspectives, public perceptions, and documentation thereof (59 words).

the value of F1 shows that the Neural Network and Bayes Point machine have the best performance, 36% and 35%, respectively, because the F1 metric value score combines the sensitivity and specificity of the considered algorithms.

An additional research study also justified the superiority of the Support Vector Machine classifier in sentiment analysis regarding public health issues. This specific research validated the accuracy of machine learning algorithms during a collection of English posts from Instagram during the pandemic 70.

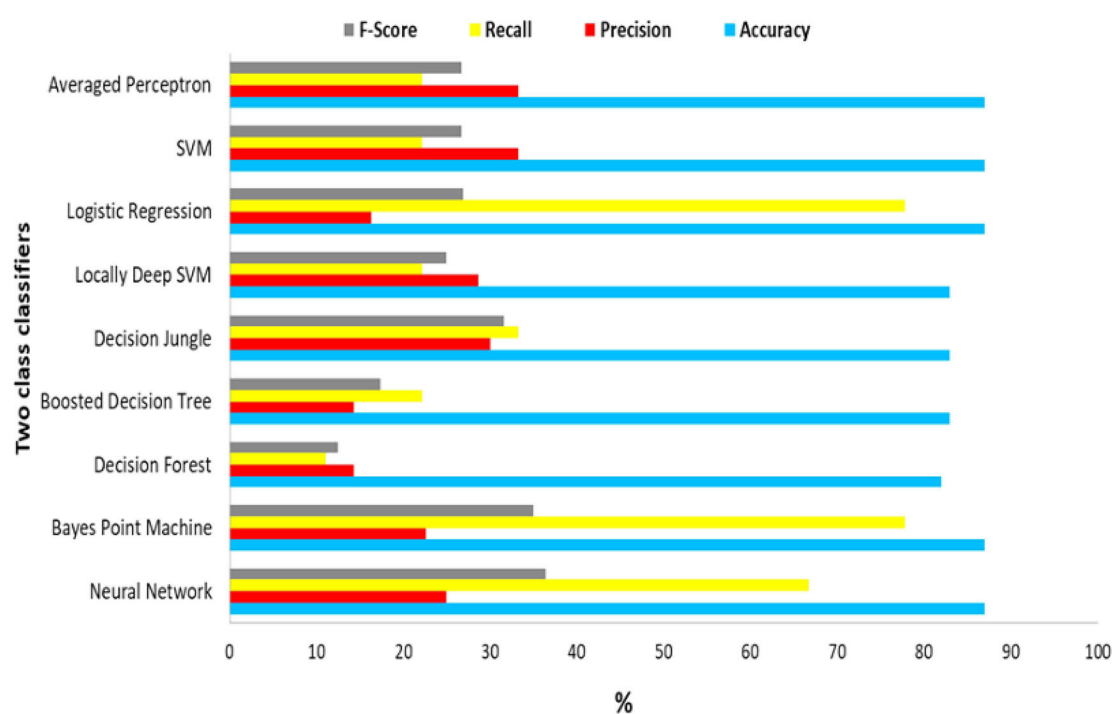
A similar study applied machine learning models for sentiment analysis. It proved that the multilayer perceptron and the support vector machine algorithm had the best evaluation score, with 76% and 74% accuracy, respectively (25).

A further study confirmed that Linear Regression, Random Forest, and Decision Tree classifiers achieved an excellent accuracy score, while the accuracy of the Support Vector Machine was 95%. After cross validation of the models used in this research, the Support Vector Machine and Random Forest model proved

sufficient (70). One more research conducted in Jordan considered Facebook posts highlighted the advanced performance of Support Vector Machine in alignment with a second algorithm, the Whale Optimization. The accuracy score ranged from 69,05% to 84,64% in various datasets (28).

In the same context, an extra research paper proposed a deep learning model for sentiment analysis and classification regarding two datasets of tweets from January 2019 to March 2020 and December 2019 until May 2020. The best accuracy results were succeeded with the Logistic Regression, which is in line with the results of our research, and the Random Forest classifiers reaching up to 75% and 81%, respectively 72.

Users' satisfaction levels toward governmental mobile applications in Saudi Arabia and respective sentiment analyses were obtained from another inquiry (71). Five machine learning classifiers were implemented: random forest (RF), bagging, support vector machine (SVM), logistic regression (LR), and naïve Bayes (NB). They revealed that the Support Vector Machine achieved



Two class classifiers	Accuracy	Precision	Recall	F-Score	Value %
Neural Network	87	25	67	36	100
Bayes Point Machine	87	23	78	35	90
Decision Forest	82	14	11	13	70
Boosted Decision Tree	83	14	22	17	50
Decision Jungle	83	30	33	32	40
Locally Deep SVM	83	29	22	25	30
Logistic Regression	87	16	78	27	20
SVM	87	33	22	27	10
Averaged Perceptron	87	33	22	27	0

FIGURE 6

Evaluation of two class classification machine learning algorithms and the performance heat map.

the best accuracy score (94.38%) implementing the SMOTE technique. In line with this study, the researchers conducted a sentiment analysis of users' opinions of mobile Apps that the Saudi Arab government introduced to combat the pandemic. They reported that the K-Nearest Neighbor and Decision Tree classifiers outperformed in terms of accuracy by 78% and 60%. Support Vector Machine and Naïve Bayes classifiers accomplished 55% and 51% accuracy scores (72, 73).

The exploitation of social network data to develop reliable early information surveillance and warning system for pandemic outbreaks resulted in a three-part integrated system. Twitter extracted data were tested for sentiment classification using various classifiers that outlined the spatiotemporal dimension in the early period of the COVID-19 outbreak. A version of the Decision Tree Classifier outperformed conventional sentiment and geolocation classification models, achieving 94.3% and 80.8% Accuracy values, respectively (74). In our sentiment analysis, Decision

Forest and Decision Jungle classifiers scored Accuracy values over 80%.

Precise sentiment classification is vital to accurate predictions on infection disease evolution. In this line, sentiment analysis at the word and document level was performed using two machine learning algorithms. Accuracy values were equivalent to ~87% and 92%, aligning with our research results since all algorithms' corresponding accuracy metrics ranged between 83% and 87% (75).

Our literature search was limited to peer-reviewed publications in English, indexed in the Scopus database in the Related Work section. The choice of databases and keywords for the literature search impacted the number of studies selected for this study. Some data source biases are inherent to social media, such as authentication issues of the user profile. Other biases that researchers could have accounted for were posts generated from bots and non-individual accounts. Also, the data set could be more extended, but extracting comments from Facebook is relatively

more complex than other platforms, although Facebook is the most popular platform in Greece. Topics of sarcasm, irony, informal expressions, humor, and slang are challenging to detect by computer programs (49, 76, 77). In addition, a large data set could improve the accuracy of our results. Another limitation of our study is the lack of an occurrence map that considers the words appearing in the word cloud (Figure 5). Such a map could provide insights into the public's perception and topics of discussion, allowing for the conceptualization and determination of the connections between sentiment, topics, and opinions on public health issues.

Despite the limitations mentioned above, this study has presented a case study conducted in Greece, which identifies key areas that should be taken into account by researchers, health professionals, health organizations, and crisis communications managers. Future initiatives will examine the performance of additional machine learning techniques. Also, a more considerable scale evaluation will be performed to provide a more complete insight into the evaluation metrics of the algorithms.

## 4. Conclusion

Considering that the COVID-19 pandemic crisis is still a major concern to governmental and public agencies, sentiment analysis can provide a better insight into the ongoing side effects in social terms, as it can predict the positive, negative, and neutral orientation of citizens. The global increase and broad use of Social Networking Sites have initiated new forms of expressing opinions and sentiments on SNS like Twitter, Facebook, etc.

This research analyzed the public perception and sentiment orientation regarding the daily surveillance reports of EODY in Greece. The respective sentiments were classified into three main comment categories: positive, neutral, and negative. Moreover, this study implemented machine learning algorithms to predict public positive or negative sentiment orientation concerning Facebook and COVID-19 data from daily surveillance reports. Furthermore, this research study covered a wide range of sentiment analysis approaches from the state of the art classifiers and compared each performance in terms of accuracy, F1 score, recall and precision. Overall, two classifiers, namely two class Neural Network and two class Bayes Point Machine, achieved high sentiment analysis accuracy and F1 score, particularly 87% and over 35%.

Based on the results, sentiment analysis and prediction models can provide critical supplementary information about the expression of sentiments during the COVID-19 pandemic. Gender and time are two factors that may determine public opinion on medical topics and especially those regarding the pandemic crisis. Overall, further research is required to advance algorithms and predicting models for sentiment and opinions to help monitor public health services and decision making processes during the pandemic.

In future work, new extended sentiment analysis could be created by implementing new classifications, considering more generated data from Greek Social Networking Sites, and keeping the current study up to date. Thus, it could extend prospective sentiment analysis using multiple models and provide additional studies to interpret and manage the outcomes.

## Author contributions

Conceptualization: CS and EG. Methodology: CS. Data curation: AT. Writing—original draft preparation: KK, CS, and CK. Writing—review and editing: EN, CT, and YK. Supervision: EC and AC. Project administration: EB and ID. All authors have read and agreed to the published version of the manuscript.

## Acknowledgments

We acknowledge the support of this work by the project Risk and Resilience Assessment Center—Prefecture of East Macedonia and Thrace—Greece (MIS 5047293), which is implemented under the Action Reinforcement of the Research and Innovation Infrastructure, funded by the Operational Programme Competitiveness, Entrepreneurship and Innovation (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpubh.2023.1191730/full#supplementary-material>

## References

- Gollust SE, Nagler RH, Fowler EF. The Emergence of in the US: a public health and political communication crisis. *J Health Polit Policy Law*. (2020) 45:967–81. doi: 10.1215/03616878-8641506
- Malecki K.MC, Keating JA, Safdar N. Crisis communication and public perception of COVID-19 risk in the era of social media. *Clin Infectious Dis*. (2021) 72:4697–702. doi: 10.1093/cid/ciaa758
- Galetsi P, Katsaliaki K, Kumar S. The medical and societal impact of big data analytics and artificial intelligence applications in combating pandemics: a review focused on COVID-19. *SocSci Med*. (2022) 301:114973. doi: 10.1016/j.socscimed.2022.114973
- Wang Y, Hao H, Platt LP. Examining risk and crisis communications of government agencies and stakeholders during early-stages of COVID-19 on Twitter. *Comp Human Behav*. (2021) 114:106568. doi: 10.1016/j.chb.2020.106568
- Quinn P. Crisis communication in public health emergencies: the limits of 'legal control' and the risks for harmful outcomes in a digital age. *Life Sci Soc Policy*. (2018) 14:4. doi: 10.1186/s40504-018-0067-0
- Huang X, Wang S, Zhang M, Hu T, Hohl A. Social media mining under the COVID-19 context: progress challenges and opportunities. *Int J App Earth Observ Geoinformation*. (2022) 113:102967. doi: 10.1016/j.jag.2022.102967
- Tsao SF, Chen H, Tisseverasinghe T, Yang H, Li L. What social media told us in the time of COVID-19: a scoping review. *Lancet Digital Health*. (2021) 3:e175–94. doi: 10.1016/S2589-7500(20)30315-0
- Kakkalou C, Nikolaou A. *Platform for Unstructured Data Analysis—Knowledge Mining From Social Media for Public Health Scenarios*. Master Thesis Aristotle. Thessaloniki: University of Thessaloniki Thessaloniki Greece. (2018).
- Yeung D. Social media as a catalyst for policy action and social change for health and well-being: viewpoint. *J Med Internet Res*. (2018) 20:e94. doi: 10.2196/jmir.8508
- Bansal S, Chowell G, Simonsen L, Vespignani A, Viboud C. Big data for infectious disease surveillance and modeling. *J Infect Dis*. (2016) 214(suppl\_4):S375–9. doi: 10.1093/infdis/jiw400
- Zhao Y, He X, Feng Z, Bost S, Prosperi M, Wu Y. Biases in using social media data for public health surveillance: a scoping review. *Int J Med Inform*. (2022) 164:04804. doi: 10.1016/j.ijmedinf.2022.104804
- Grajales III FJ, Sheps S, Ho K, Novak-Lauscher H, Eysenbach G. Social media—A review and tutorial of applications in medicine and health care. *J Med Internet Res*. (2014) 16:e13. doi: 10.2196/jmir.2912
- Frey E, Bonfiglioli C, Brunner M, Frawley J. Parents' use of social media as a health information source for their children: a scoping review. *Acad Pediatr*. (2022) 22:526–39. doi: 10.1016/j.acap.2021.12.006
- Kanita Karadzović-Hadžabiđić Rialda Spahić Emin Tahirović. Evaluation of IBM Watson Natural Language Processing Service to predict influenza-like illness outbreaks from Twitter data. *Periodicals of Engineering and Natural Sciences Original Research* 10 1 January 2022 pp.122–137. (2022).
- EODY. Available online at: <https://eody.gov.gr/en/> (accessed July 12, 2022).
- Stefanis C, Giorgi E, Kalentzis K, Tseleponis A, Tsigalou C, Nena E, et al. Assessing worldwide research activity on ICT in climate change using Scopus database: a bibliometric analysis. *Front Environ Sci*. (2022) 10:198. doi: 10.3389/fenvs.2022.868197
- Stefanis C, Stavropoulou E, Giorgi E, Voidarou CC, Constantinidis TC, Vrioni G, et al. Honey's antioxidant and antimicrobial properties: a bibliometric study. *Antioxidants*. (2023) 12:414. doi: 10.3390/antiox12020414
- Islam Nowman Anwar Aamir and Rehman IkramA proposed framework for developing user-centred mobile healthcare applications for the biggest annual mass gathering (Hajj) post COVID-19. In: *34<sup>th</sup> British Human Computer Interaction Conference* 19–21 Jul 2021 London UK. (2021).
- Jain VK, Kumar S. Effective surveillance and predictive mapping of mosquito-borne diseases using social media. *J Comp Sci*. (2018) 25:406–415. doi: 10.1016/j.jocs.2017.07.003
- Loukis E. Citizen-Sourcing for Public Policy Making. *Theor Found Methods Eval*. (2018) 3:8. doi: 10.1007/978-3-319-61762-6\_8
- Walter D, Bohmer M, Reiter S, Krause G, Wichmann O. Risk perception and information-seeking behavior during the 2009/10 influenza A(H1N1)pdm09 pandemic in Germany. *Euro Surveill*. (2012) 17:20131. doi: 10.2807/ese.17.13.20131-en
- Simões D, Ehsani S, Stanojevic M, Shubladze N, Kalmambetova G, Paredes R, et al. Integrated use of laboratory services for multiple infectious diseases in the WHO European Region during the COVID-19 pandemic and beyond. *Euro Surveill*. (2022) 27:2100930. doi: 10.2807/1560-7917.ES.2022.27.29.2100930
- Samuel J, Ali GGMN, Rahman MM, Esawi E, Samuel Y. COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification. *Information*. (2020) 11:314. doi: 10.3390/info11060314
- Wrycza S, Maślankowski J. Social media users' opinions on remote work during the COVID-19 Pandemic. *Them Senti Anal Inform Sys Manag*. (2020) 37:288–97. doi: 10.1080/10580530.2020.1820631
- Babić K, Petrović M, Beliga S, Martinčić-Ipšić S, Matešić M, Meštrović A. Characterisation of COVID-19-related tweets in the Croatian language: framework based on the Cro-CoV-cseBERT model. *Appl Sci*. (2021) 11:10442. doi: 10.3390/app112110442
- Sattar NS, Arifuzzaman S. COVID-19 Vaccination awareness and aftermath: public sentiment analysis on twitter data and vaccinated population prediction in the USA. *Appl Sci*. (2021) 11:6128. doi: 10.3390/app11136128
- Nemes L, Kiss A. Information extraction and named entity recognition supported social media sentiment analysis during the COVID-19 pandemic. *Appl Sci*. (2021) 11:1017. doi: 10.3390/app112211017
- Obiedat R, Harfoushi O, Qaddoura R, Al-Qaisi L, Al-Zoubi AM. An evolutionary-based sentiment analysis approach for enhancing government decisions during COVID-19 pandemic: the case of Jordan. *Appl Sci*. (2021) 11:9080. doi: 10.3390/app11199080
- Raptopoulos I. *Sentiment Analysis in Social Networks*. Bachelor Thesis. Patra: University of Patras Greece. (2019).
- Liodakis M, Aleksandrakis E. *Sentiment Analysis of Greek Text Using Machine Learning Algorithms*. Bachelor Thesis. Pirais: University of Pirais Greece. (2017).
- Kardakis S. *Machine Learning Techniques for Sentiment Analysis and Emotion Recognition in Natural Language*. Patra: University of Patras Greece. (2019).
- Birbili K. *Analysis of Sentiment in Social Networks regarding Economic Measures in Greece*. Master Thesis. Patra: University of Patras Greece (2016.)
- Pantoglou D. *Sentiment Analysis of Greek texts From Social Media Using Statistical Learning Algorithms*. Master Thesis Aristotle. Thessaloniki: University of Thessaloniki Greece. (2019).
- Kostidis C. *Techniques for Attribute Based Sentiment Analysis on Social Networks*. Bachelor Thesis Aristotle. Thessaloniki: University of Thessaloniki Greece. (2017).
- Mitsopoulou E. *Machine Learning-Based Sentiment Analysis of Twitter Data*. Master Thesis Aristotle. Thessaloniki: University of Thessaloniki Greece. (2020).
- Papaioannoy K. *Automated Retrieval and Processing of Scientific Literature in Order to Evaluate Medical Hypotheses*. Master Thesis Aristotle. Thessaloniki: University of Thessaloniki Greece. (2018).
- Walsh J, Dwumfour C, Cave J. Spontaneously generated online patient experience data - how and why is it being used in health research: an umbrella scoping review. *BMC Med Res Methodol*. (2022) 22:139. doi: 10.1186/s12874-022-01610-z
- Pilipiec P, Samsten I, Bota A. Surveillance of communicable diseases using social media: a systematic review. *PLoS ONE*. (2023) 18:e0282101. doi: 10.1371/journal.pone.0282101
- Sinnenberg I, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: a systematic review. *Am J Public Health*. (2017) 107:e1–8. doi: 10.2105/AJPH.2016.303512
- Sun H, Zhang Y, Gao G, Wu D. Internet search data with spatiotemporal analysis in infectious disease surveillance: challenges and perspectives. *Front Public Health*. (2022) 10:958835. doi: 10.3389/fpubh.2022.958835
- Salas-Zarate MP, Medina-Moreira J, Lagos-Ortiz K, Luna-Aveiga H, Rodríguez-García M, García RV. Sentiment analysis on tweets about diabetes: an aspect-level approach. *Comp Math Methods Med*. (2017) 9:5140631. doi: 10.1155/2017/5140631
- Park S, Woo J. Gender classification using sentiment analysis and deep learning in a health web forum. *Appl Sci*. (2019) 9:1249. doi: 10.3390/app9061249
- Xu Q, Chang V, Jayne C. A systematic review of social media-based sentiment analysis: emerging trends and challenges. *Dec Anal J*. (2022) 3:100073. doi: 10.1016/j.dajour.2022.100073
- Karim M, Missen MMS, Umer M, Sadiq S, Mohamed A, Ashraf I. Citation context analysis using combined feature embedding and deep convolutional neural network model. *App Sci*. (2022) 12:203. doi: 10.3390/app12063203
- Hall K, Chang V, Jayne C. A review on natural language processing models for COVID-19 research. *Healthcare Anal*. (2022) 2:2772–4425. doi: 10.1016/j.health.2022.100078
- Alexandros B. *Sentiment Analysis on Streams of Twitter Data*. Patra: Dissertation University of Patras Greece. (2016).
- MacKay M, Cimino A, Yousef Inaghani S, McWhirter JE, Dara R, Papadopoulos A. Canadian COVID-19 crisis communication on twitter: mixed methods research examining tweets from government politicians and public health for crisis communication guiding principles and tweet engagement. *Int J Environ Res Public Health*. (2022) 19:6954. doi: 10.3390/ijerph19116954



48. Liotiri E. Sentiment Analysis using machine learning techniques and neural nets in Twitter. Master Thesis Aristotle University of Thessaloniki Thessaloniki Greece. (2019).
49. Shalev-Shwartz S, Ben-David S. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge: Cambridge University Press (2014).
50. Fushiki T. Estimation of prediction error by using K-fold cross-validation. *Stat Comput.* (2021) 21:137–46. doi: 10.1007/s11222-009-9153-8
51. Nedel'ko VM. Statistical Fitting Criterion on the Basis of Cross-Validation Estimation. *Pattern Recognit. Image Ana.* (2018) 28:510–15. doi: 10.1134/S1054661818030148
52. Zhao P, Liu F, Zhuang X. Speech sentiment analysis using hierarchical conformer networks. *Appl Sci.* (2022) 12:8076. doi: 10.3390/app12168076
53. Jnoub N, Al Machot F, Klas W. A Domain-independent classification model for sentiment analysis using neural models. *Appl Sci.* (2020) 10:6221. doi: 10.3390/app10186221
54. AlGhamdi N, Khatoun S, Alshamari M. Multi-aspect oriented sentiment classification: prior knowledge topic modelling and ensemble learning classifier approach. *Appl Sci.* (2022) 12:4066. doi: 10.3390/app12084066
55. Sesagiri Raamkumar A, Tan SG, Wee HL. Measuring the outreach efforts of public health authorities and the public response on Facebook during the COVID-19 pandemic in early 2020: cross-country comparison. *J Med Internet Res.* (2020) 22:e19334. doi: 10.2196/19334
56. Hussain A, Tahir A, Hussain Z. Artificial intelligence-enabled analysis of public attitudes on Facebook and Twitter toward COVID-19 vaccines in the United Kingdom and the United States: observational study. *J Med Internet Res.* (2021) 23:e26627. doi: 10.2196/26627
57. Wang T, Lu K, Chow K, Zhu Q. COVID-19 sensing: negative sentiment analysis on social media in China via BERT model. *IEEE Access.* (2020) 8:138162–9. doi: 10.1109/ACCESS.2020.3012595
58. MacKay M, Colangeli T, Gillis D, McWhirter J, Papadopoulos A. Examining social media crisis communication during early COVID-19 from public health and news media for quality content and corresponding public sentiment. *Int J Environ Res Public Health.* (2021) 18:7986. doi: 10.3390/ijerph18157986
59. He S, Li D, Liu CH, Xiong Y, Liu D, Feng J, et al. Crisis communication in the WHO COVID-19 press conferences: a retrospective analysis. *PLoS One.* (2023) 18:e0282855. doi: 10.1371/journal.pone.0282855
60. Enria L, Waterlow N, Rogers NT, Brindle H, Lal S, Eggo RM, et al. Trust and transparency in times of crisis: results from an online survey during the first wave (April 2020) of the COVID-19 epidemic in the UK. *PLoS One.* (2021) 16:e0239247. doi: 10.1371/journal.pone.0239247
61. Larson HJ, Jarrett C, Eckersberger E, Smith DMD, Paterson P. Understanding vaccine hesitancy around vaccines and vaccination from a global perspective: a systematic review of published literature 2007–2012. *Vaccine.* (2014) 32:2150–9. doi: 10.1016/j.vaccine.2014.01.081
62. Bish A, Yardley L, Nicoll A, Michie S. Factors associated with uptake of vaccination against pandemic influenza: a systematic review. *Vaccine.* (2011) 29:6472–84. doi: 10.1016/j.vaccine.2011.06.107
63. Lovari A, Martino V, Righetti N. Blurred shots: investigating the information crisis around vaccination in Italy. *Am Behav Sci.* (2020) 5:000276422091024. doi: 10.1177/0002764220910245
64. Dowd JB, Andriano L, Brazel DM, Rotondi V, Block P, Ding X, et al. Demographic science aids in understanding the spread and fatality rates of COVID-19. *Proc Natl Acad Sci USA.* (2020) 117:9696–8. doi: 10.1073/pnas.2004911117
65. Available online at: <https://royalsociety.org/-/media/policy/projects/set-c/set-c-vaccine-deployment.pdf> Block (accessed 21/06/2023).
66. Basant, Vaishnavi S, Priyanka H, Vinita H, Sharma A. The COVID-19 outbreak: social media sentiment analysis of public reactions with a multidimensional perspective. *Cyber-Physical Sys.* (2022) 4:117–38. doi: 10.1016/B978-0-12-824557-6.00013-3
67. Xue J, Chen, J, Hu R, Chen C, Zheng C, et al. Twitter discussions and emotions about the COVID-19 pandemic: machine learning approach. *J Med Internet Res.* (2020) 25:22:e20550. doi: 10.2196/20550
68. Lwin M, Lu J, Sheldenkar A, Schulz P, Shin W, Gupta R, et al. Global sentiments surrounding the COVID-19 pandemic on Twitter: analysis of Twitter trends. *JMIR Public Health Surveill.* (2020) 6:e19447. doi: 10.2196/19447
69. Sánchez-Garcés J, López-Gonzales JL, Palacio-Farfán M, Coronel-Sacón V, Fernéy-Teheran Y, Peña-Pineda J, et al. exploratory analysis of fundamental spiritual support factors to a positive attitude in patients with COVID-19 using natural-language processing algorithms. *Appl Sci.* (2021) 11:9524.
70. Amanatidis D, Mylona I, Kamenidou I, Mamalis S, Stavrianea A. Mining textual and imagery instagram data during the COVID-19 pandemic. *Appl Sci.* (2021) 11:4281. doi: 10.3390/app11094281
71. Mujahid M, Lee E, Rustam F, Washington PB, Ullah S, Reshi AA, et al. Sentiment analysis and topic modeling on tweets about online education during COVID-19. *Appl Sci.* (2021) 11:8438. doi: 10.3390/app11188438
72. Chakraborty K, Bhatia S, Bhattacharyya S, Platos J, Bag R, Hassanien AE. Sentiment analysis of COVID-19 tweets by deep learning classifiers—A study to show how popularity is affecting accuracy in social media. *Appl Soft Comput J.* (2020) 97:106754. doi: 10.1016/j.asoc.2020.106754
73. Hadwan M, Al-Sarem M, Saeed F, Al-Hagery MA. An Improved sentiment classification approach for measuring user satisfaction toward governmental services' mobile apps using machine learning methods with feature engineering and SMOTE technique. *Appl Sci.* (2022) 12:5547. doi: 10.3390/app12115547
74. Hadwan M, Al-Hagery M, Al-Sarem M, Saeed F. Arabic sentiment analysis of users' opinions of government mobile applications. *Comput Mater Contin.* (2022) 72:4675–89. doi: 10.32604/cmc.2022.027311
75. Gamal N, Ghoniemy S, Faheem HM, Seada NA. Sentiment-based spatiotemporal prediction framework for pandemic outbreaks awareness using social networks data classification. *IEEE Access.* (2022) 10:76434–69. doi: 10.1109/ACCESS.2022.3192417
76. Shan S, Yan Q, Wei Y. Infectious or recovered? Optimizing the infectious disease detection process for epidemic control and prevention based on social media. *Int J Environ Res Public Health.* (2020) 17:6853. doi: 10.3390/ijerph17186853
77. Spatiotis N, Mporas I, Paraskevas D, Perikos I. Sentiment analysis for the Greek language. In *Proceedings of the 20th Pan-Hellenic Conference on Informatics.* (2016) 10:1–4. doi: 10.1145/3003733.3003769