Systematic Review and Meta-Analysis

Medicine

OPEN

COVID-19 studies involving machine learning methods

A bibliometric study

Arzu Baygül Eden, PhD[a,*], Alev Bakır Kayı, PhD[b], Mustafa Genco Erdem, PhD, MD[c], Mehmet Demirci, PhD[d]

Abstract

Background: Machine learning (ML) and artificial intelligence (AI) techniques are gaining popularity as effective tools for coronavirus disease of 2019 (COVID-19) research. These strategies can be used in diagnosis, prognosis, therapy, and public health management. Bibliometric analysis quantifies the quality and impact of scholarly publications. ML in COVID-19 research is the focus of this bibliometric analysis.

Methods: A comprehensive literature study found ML-based COVID-19 research. Web of Science (WoS) was used for the study. The searches included "machine learning," "artificial intelligence," and COVID-19. To find all relevant studies, 2 reviewers searched independently. The network visualization was analyzed using VOSviewer 1.6.19.

Results: In the WoS Core, the average citation count was 13.6 ± 41.3. The main research areas were computer science, engineering, and science and technology. According to document count, Tao Huang wrote 14 studies, Fadi Al-Turjman wrote 11, and Imran Ashraf wrote 11. The US, China, and India produced the most studies and citations. The most prolific research institutions were Harvard Medical School, Huazhong University of Science and Technology, and King Abdulaziz University. In contrast, Nankai University, Oxford, and Imperial College London were the most mentioned organizations, reflecting their significant research contributions. First, "Covid-19" appeared 1983 times, followed by "machine learning" and "deep learning." The US Department of Health and Human Services funded this topic most heavily. Huang Tao, Feng Kaiyan, and Ashraf Imran pioneered bibliographic coupling.

Conclusion: This study provides useful insights for academics and clinicians studying COVID-19 using ML. Through bibliometric data analysis, scholars can learn about highly recognized and productive authors and countries, as well as the publications with the most citations and keywords. New data and methodologies from the pandemic are expected to advance ML and AI modeling. It is crucial to recognize that these studies will pioneer this subject.

Abbreviations: AI = artificial intelligence, COVID-19 = coronavirus disease of 2019, DL = deep learning, LDH = lactic dehydrogenase, ML = machine learning, SEIR = susceptible-exposed-infectious-removed, USA = United States of America, WoS = web of science.

Keywords: artificial intelligence, bibliometric analysis, COVID-19, machine learning

1. Introduction

The emergence of the coronavirus illness (COVID-19) in Wuhan marked the beginning of a global health crisis. On March 11, 2020, the World Health Organization officially declared it as a pandemic. Since then, the impact of this disease has been significant, affecting numerous individuals worldwide. As of March 2023, there have been over 500 million confirmed cases and more than 6 million documented deaths globally.[1] COVID-19 has exerted a significant impact on the scientific literature, alongside its repercussions on society. The global pandemic has resulted in an exceptional surge in research activity, leading to the publication of a substantial number of publications across several academic fields, encompassing medicine and health informatics among others.

Artificial intelligence (AI) can be conceptualized as a paradigm wherein computers and state-of-the-art technology strive to replicate human-like intelligent behavior and critical thinking by means of expeditious data processing. The phrase "artificial intelligence" was initially used by John McCarthy in 1956,

The authors have no funding and conflicts of interest to disclose.

The datasets generated during and/or analyzed during the current study are publicly available.

[a] Koç University, School of Medicine, Department of Biostatistics, Istanbul, Turkey, [b] Istanbul University, Institute of Child Health, Department of Social Pediatrics, Istanbul, Türkiye, [c] Department of Internal Medicine, Faculty of Medicine, Beykent University, Istanbul, Turkey, [d] Department of Medical Microbiology, Faculty of Medicine, Kirklareli University, Kirklareli, Turkey.

**Correspondence: Arzu Baygül Eden, Koç University, School of Medicine, Department of Biostatistics, Istanbul 34010, Turkey (e-mail: abaygul@ku.edu.tr).*

Copyright © 2023 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial License 4.0 (CCBY-NC), where it is permissible to download, share, remix, transform, and buildup the work provided it is properly cited. The work cannot be used commercially without permission from the journal.

How to cite this article: Baygül Eden A, Bakır Kayı A, Erdem MG, Demirci M. COVID-19 studies involving machine learning methods: A bibliometric study. Medicine 2023;102:43(e35564).

Received: 21 August 2023 / Received in final form: 16 September 2023 / Accepted: 18 September 2023

http://dx.doi.org/10.1097/MD.0000000000035564

defining it as "the scientific and engineering discipline concerned with the creation of intelligent machines."[1]

Machine learning (ML) is a form of AI that has revolutionized the twenty-first century. Advances in fundamental architectures and algorithms, as well as the growth in the size of datasets, have led to an increase in computer competence in different areas.[2]

Deep Learning (DL), a branch of ML and AI, is today recognized as the core technology of the Industrial Revolution. DL technology is acknowledged as a pivotal subject within the domains of ML, AI, data science, and analytics, primarily owing to its capacity to acquire knowledge from data. Given its capacity to yield significant and prompt outcomes in various classification and regression problems, as well as the rapid analysis of datasets, it is actively pursued by numerous global giants like Google and Microsoft. From a research perspective, DL is categorized as a subfield of ML and AI, serving as an AI function that emulates the data processing capabilities of the human brain.[3]

AI is a significant scientific discipline that focuses on the analysis of intricate datasets and the execution of estimations. ML is a significant component of AI that involves the construction of models and the generation of predictions. This process entails the utilization of a subset of a dataset for training purposes, while another subset is employed for testing. AI and ML methods offer a valuable chance to assess the prognosis of COVID-19 and generate predictions at both the individual and societal levels.[4]

AI models typically work with extensive feature datasets, often comprising thousands of data points. These data points form intricate patterns that are routinely observed, interpreted, and documented. The entire modeling process is often referred to as a "black box," with the inner workings from input to output not always transparent. In AI modeling, the primary objective often revolves around generating predictions for specific outcomes. Unlike traditional statistical analyses, ML models aim to predict the correct answers and can enhance their performance through self-assessment and learning from their own errors.[5]

The utilization of ML and AI techniques has become increasingly prevalent in the realm of COVID-19 research. These methods have demonstrated considerable potential in various domains, including but not limited to pulmonology, pharmacology, and biochemistry. Their application spans across multiple areas such as diagnosis, prognosis, therapy, and public health management.[6–10]

The inclusion of a literature review holds significant value within the realm of academic research as it serves as a crucial means of gathering pertinent information pertaining to a particular subject matter.[11] Systematic reviews offer a rigorous, empirical, and transparent approach to accessing a large number of published studies.[12] The transition from performing systematic reviews to employing bibliometric analysis signifies a paradigm shift in our comprehension and integration of research. Systematic reviews are comprehensive and rigorous evaluations of the available literature aimed at addressing specific research inquiries. In contrast, bibliometric analysis is a quantitative methodology that concentrates on the assessment of publishing patterns and citation metrics. Within this particular review, researchers have transitioned from engaging in critical analysis of specific works to exploring the broader scope of scholarly output.

Bibliometric analysis is a quantitative approach employed to evaluate the caliber of scholarly publications and examine their influence within a certain domain or subject matter.[13,14] The utilization of bibliometric analysis offers significant contributions in understanding the present condition of research, identifying the most influential researchers and institutions, discerning essential themes and trends, and identifying potential gaps and opportunities for future research. This study conducted a bibliometric analysis of the ML techniques employed COVID-19 research, with specific attention to the following research inquiries:

1. What is the overall publication trend regarding ML methods in COVID-19 studies?
2. Who are the most influential authors and institutions in this field?
3. What are the key research topics and trends associated with machine-learning methods in COVID-19 studies?
4. What are the most important countries in this field?
5. What are the most relevant journals in the field?
6. What is the most prolific pattern of authorship in the field?
7. What are the global most cited documents?
8. What are the most influential funding agencies in this field?
9. What are the most common author keywords?
10. What are the most collaborative countries in this field?
11. What are the major themes of research in this field?

In order to address these inquiries, a bibliometric methodology will be employed to examine an extensive collection of scholarly articles pertaining to the utilization of ML techniques in research investigations related to COVID-19.

The present study offers a thorough examination of the existing body of research in this particular topic and highlights potential opportunities for future research endeavors and collaborative efforts. The outcomes of this investigation will be of significance to scholars, professionals, policymakers, and other individuals with a stake in COVID-19 research and the utilization of ML techniques.

2. Material and methods

2.1. Search strategy

A comprehensive systematic evaluation was undertaken to locate and analyze the current pool of published material pertaining to the utilization of ML techniques in the context of COVID-19.

A comprehensive search of the scientific literature was conducted to identify studies using ML methods in COVID-19 research. The search was performed using Web of Science (WoS) on March 30, 2023. The search was limited to studies published between December 1, 2019, and December 31, 2022. The terms searched for included "Machine Learning," "Artificial Intelligence," and "COVID-19," using the following search term: TS=(("Coronavirus" OR "COVID19" OR "COVID-19" OR "Coronavirus disease 2019" OR "SARS-COV2" OR "2019 nCOV" OR "nCOV-2019") AND ("Machine Learning" OR "Artificial Intelligence")). The search was conducted independently by 2 reviewers to ensure that all relevant studies were identified. The data were imported into plain text. The plain text databases were concatenated using the VOSviewer software.

At the first screening in WoS 7102 studies about COVID-19 and ML were retrieved. Authors excluded review articles, data papers, book chapters, and non-English studies (69 studies). Among the 7033 studies, 3474 were excluded. 3559 studies were added to the final analysis (Fig. 1).

In order to mitigate bias, the review was conducted separately by 2 researchers, namely A.B.E. and A.B.K. Discrepancies in the decisions for inclusion made by the reviewers were compared. Consensus among the reviewers led to the resolution of all disagreements.

2.2. Inclusion and exclusion criteria

Inclusion criteria:

- Studies about COVID-19 that included ML methods.
- Articles and early access studies.

Exclusion criteria:

- Studies about COVID-19 but without a ML method, or vice versa.
- Book chapters, review articles, data papers.
- Studies are not in English.

2.3. Data extraction and analysis

The relevant details of each study, including the authors, title, source, times cited count, sponsors, abstracts, addresses, and document type information, were documented. The VOSviewer software version 1.6.19[15] was utilized to do network visualization analysis. The analysis of the most prolific journals was conducted using Microsoft Excel 2021. There exist further programs that are tailored specifically for the purpose of bibliometric analysis. Bibliometrix is a command-line interface software, whereas VosViewer is a user-friendly tool including a graphical user interface with buttons and windows, enabling the generation of visually appealing images.[16] When conducting a comparison between CiteSpace and VosViewer, it can be observed that VosViewer offers superior graphical clarity and user-friendly features. However, it should be noted that CiteSpace offers certain advantages in the field of network analysis, specifically in relation to cluster nodes.[17] When considering the necessity of conducting our investigation, it is evident that VosViewer would be the most advantageous software application.

The databases most commonly utilized in bibliometric analyses were WoS, Scopus, and PubMed. There was a significant overlap in the information indexed in both WoS and Scopus.[18] The primary concentration of PubMed is on the fields of medicine and biomedical sciences.[19] In our research, we aimed to incorporate works that are indirectly relevant to the field of medicine.

The WoS database contains scholarly information dating back to 1900, while Scopus mostly includes publications published from 1966 forward.[19] Given that our analysis spans the time frame of 2019–2023, there is no need for concern regarding any discrepancies in time coverage across Scopus, WoS, and PubMed. Based on the provided parameters, it appears that WoS is the most efficacious database for our research.

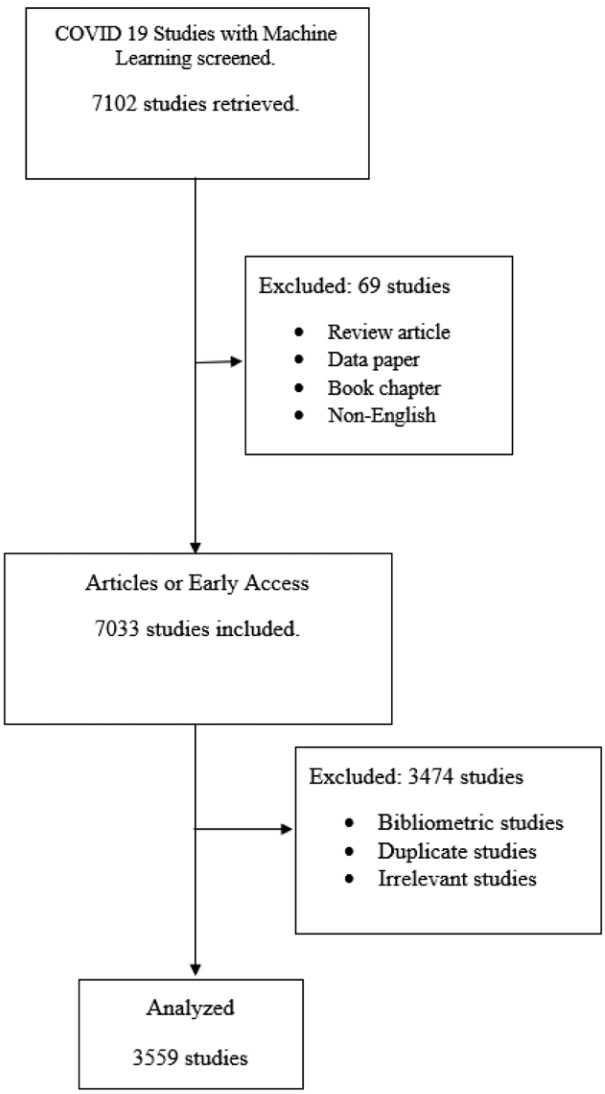


Figure 1. Flowchart of the study.

2.4. Ethical consideration

The study did not necessitate ethical approval as it solely involved the examination of preexisting data that had been previously published.

3. Results

The search yielded a total of 7102 studies. The study excluded non-English studies, bibliometric studies, review studies, book chapters, and conference papers, and only included published publications and early access articles. The final analysis incorporated a total of 3559 studies, as shown in Figure 1. Among these, 149 studies, accounting for 4.2% of the total, were found to have early access.

Since the emergence of the COVID-19 pandemic in 2019, an extensive amount of academic articles have been published. Particularly, in the year 2020, a total of 334 studies were released. This number increased significantly in 2021, with 1265 studies being published. In the subsequent year of 2022, the number of published studies further rose to 1485. As of March 30, 2023, an additional 475 studies have been published, indicating ongoing research efforts in this field.

The bibliometric study yielded insights pertaining to several aspects of research, including research areas, citation counts, authorship, countries, organizations and keywords.

The average number of citations was found to be 13.6 ± 41.3 in the WoS Core database and 13.8 ± 42.5 across all indexes.

3.1. Bibliometric analysis of research areas

The research field that gained the highest level of popularity in ML and COVID-19 studies was "computer science," accounting for 29.8% of the total. Engineering, science, and technology ranked second and third, respectively (Fig. 2).

3.2. Bibliometric analysis of authors

19.622 authors—an average of 5.51 authors per article—were responsible for the 3559 documents. The productivity of authors was evaluated according to the number of documents and citations (Fig. 3).

According to the document number, the top 3 authors were Tao Huang with 14 studies and Fadi Al-Turjman and Imran Ashraf with 11 studies.

The top 3 authors were Rashid Mazhar, Muhammad E. H. Chowdhury with 316 citations, and Tawsifur Rahman with 601, 316, and 181 citations, respectively.

The top 10 authors, who published the most documents and had the most citations do not overlap. It may be possible, because some authors can produce many studies with low citations, however some authors' have fewer studies with high citations.

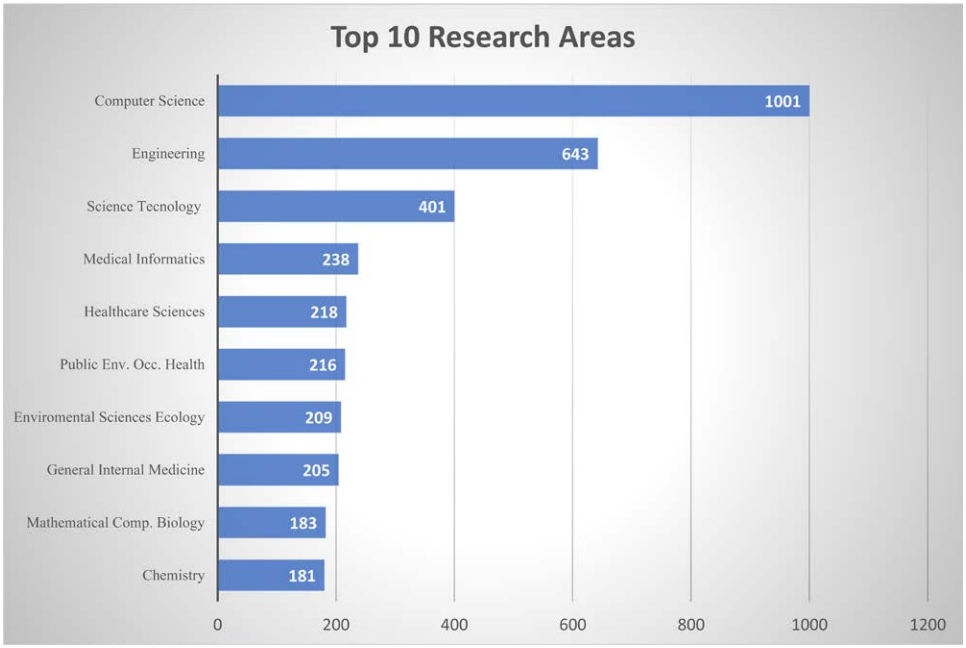


Figure 2. Top 10 research areas.

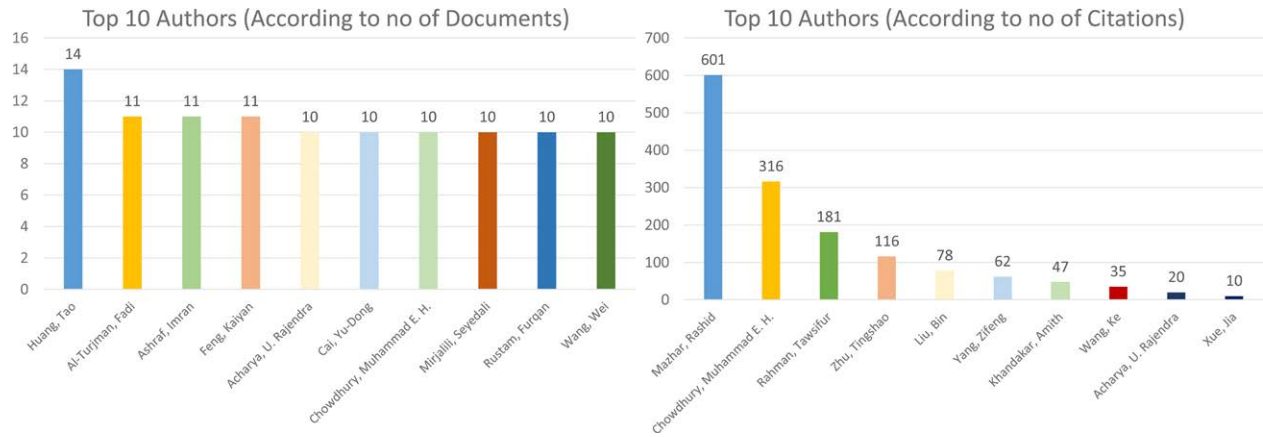


Figure 3. Top 10 authors.

A network visualization map was constructed for authors who had at least 1 study and the highest link strength (Fig. 4); 237 authors were included. Among these authors, 1830 links and 15 clusters were found, and the total link strength was 2325.

3.3. Bibliometric analysis of countries

From the bibliometric analysis, 119 countries contributed to the ML/COVID-19 literature. The distribution of the top ten most productive countries is shown in Figure 5 by document number and citations.

The most productive country for document number was the United States of America (USA), with 931 articles; the second most productive country was China (n: 546); and India (n: 527) was the third country. For citations, the first country was the USA (n:15311), the second country was China (9673 citations), and India had 7105 citations (Fig. 5).

A network visualization map for the authors was constructed for countries with at least 1 study. In total, 113 countries were included in this study. Among these authors, 1422 links and 11 clusters were identified, with a total link strength of 5002 (Fig. 6).

3.4. Bibliometric analysis of organizations

Of the 5714 organizations that produced at least 1 study on COVID-19 involving ML topics, 998 organizations with the highest link strength were included in the network visualization map. The node weights based on the number of documents, 25 clusters, 9122 links, and 10,783 total link strengths were constructed in the network visualization map (Fig. 7).

The top 3 organizations by document number were Harvard Medical School with 50 documents, Huazhong University of Science and Technology with 48 documents, and King Abdulaziz University with 47 documents. The order was changed when taking "citations" as an index. Nankai University, the University of Oxford, and Imperial College London were the top 3 universities, with 722, 150, and 54 citations, respectively (Fig. 8).

3.5. Bibliometric analysis of keywords

The minimum number of keyword occurrences was set to 5. Of the 9406 keywords, 748 met the threshold.

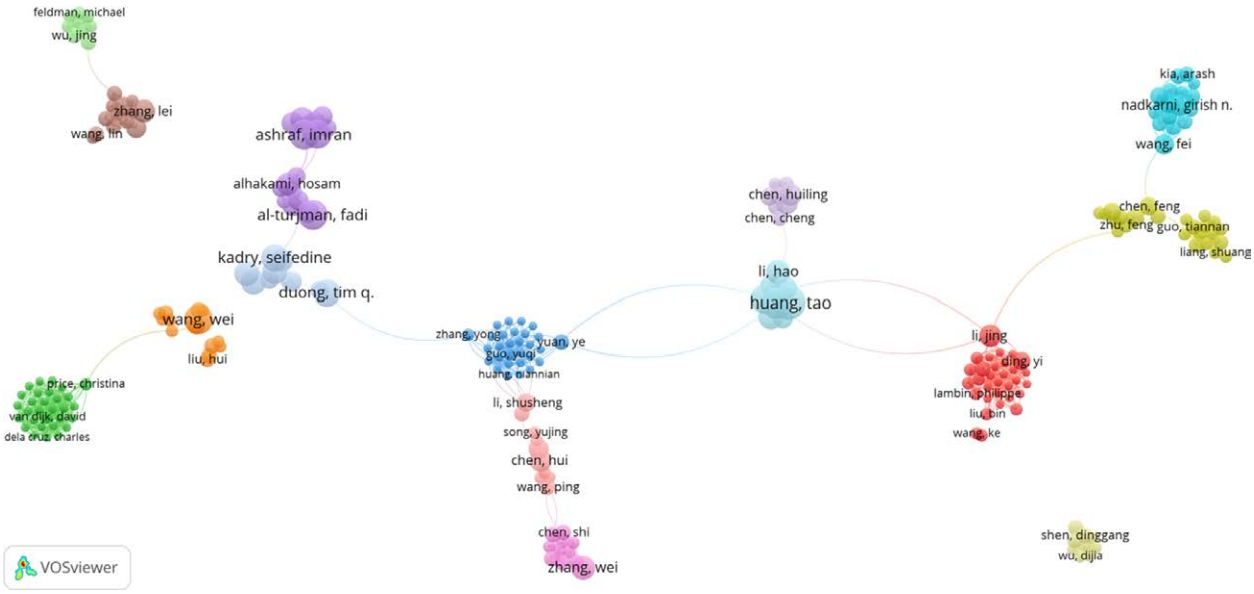


Figure 4. Network visualization map of authors.

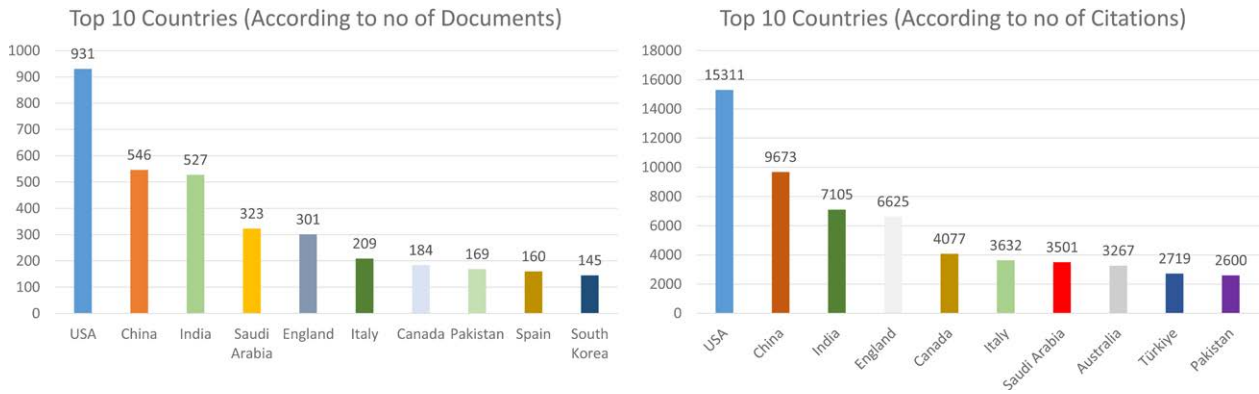


Figure 5. Top 10 countries.

The keyword "Covid-19" was in first place with 1983 occurrences, followed by "machine learning" and "deep learning" (Fig. 9). The classification, prediction, and diagnosis keywords for COVID-19 were identified in the top ten. Among the 10 most identified keywords, the keywords "COVID-19," "SARS-CoV-2," "coronavirus," and "diagnosis" were classified as clinically related keywords, while the other 6 keywords were classified as statistically related keywords (Fig. 10).

3.6. Articles

The most cited articles were "The Impact of the COVID-19 Epidemic Declaration on Psychological Consequences: A Study on Active Weibo Users" (L.S.J., W.Y.L., X.J., Z.N., Z.T.S.), "Modified SEIR and AI prediction of the epidemic trend of COVID-19 in China under public health interventions" (Y.Z.F., Z.Z.Q., W.K., W.K., W.S.S., L.W.H., Z.M., L.P., C.X.D., G.Z.Q., M.Z.T.), "Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil" (F.N.R., M.T.A., W.C., C.I.M., C.D.D., M.S., C.M.A.E., S.F.C., H.I., M.J.T., H.R.J.G., F.L.A.M., R.M.S., dJ.J.G., A.P.S., C.T.M., F.G.M., S.C.A.M., M.E.R., P.R.H.M., P.P.S., K.M.U., G.N., C.C.D., H.H., S.W.M., R.E.C., dS.L.M., dP.M.C., A.L.J.T., M.F.S.V., dL.A.B., S.J.D., Z.D.A.G., F.A.C.D., S.R.P., L.D.J., W.P.G.T., S.H.M., dS.A.L.P., V.M.S., D.C.V.S., F.R.M.F., dS.H.M., A.R.S., P.-M.J.L.P., N.B., H.J.S., M.M., M.X., C.H., S.R., V.M., G.A., P.C.A., N.V.H., S.M.A., B.T.A., P.S.L.K., W.C.H., R.O., F.N.M., D.C., L.N.J., L.P., R.A., F.N.A., C.M.D.S.S., P.O.G., F.S., B.S., S.E.C.), "Can AI help in screening viral and COVID-19 pneumonia?" (C.M.E.H., R.T., K.A., M.R., A.K.M., M.Z.B., I.K.R., K.M.S., I.A., E.N.A., R.M.B.I., I.M.T.), "An interpretable mortality prediction model for COVID-19 patients" (Y.L., Z.H.T., G.J., X.Y., W.M.L., G.Y.Q., S.C., T.X.C., J.L., Z.M.Y., H.X., X.Y., H.S.F., T.X., H.N.N., J.B., C.C., Z.Y., L.A.L., M.L., J.J.Y., C.Z.G., L.S.S., X.H., Y.Y.)[20–24] (Table 1).

3.6.1. The impact of COVID-19 epidemic declaration on psychological consequences: A study on active Weibo users. The study intends to look into how COVID-19 affects people psychologically. It examines Weibo postings from 17,865 active users using Online Ecological Recognition and ML forecasting models. Word frequency, emotional markers (such as anxiety, depression, indignation, and Oxford happiness), and cognitive indicators (such as social risk assessment and life satisfaction) are all included in the analysis. In order to compare pre- and post-COVID-19 declaration data (before and after January 20, 2020), sentiment analysis and paired sample t-tests were used. The results show a drop in good emotions (Oxford happiness) and life satisfaction combined

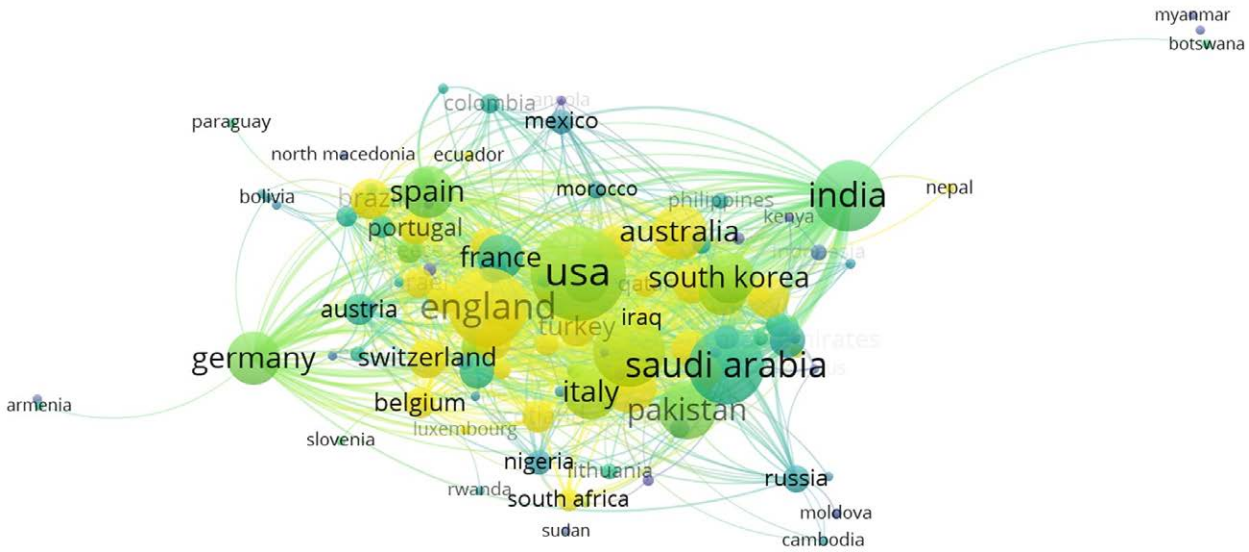


Figure 6. Network visualization map of countries.

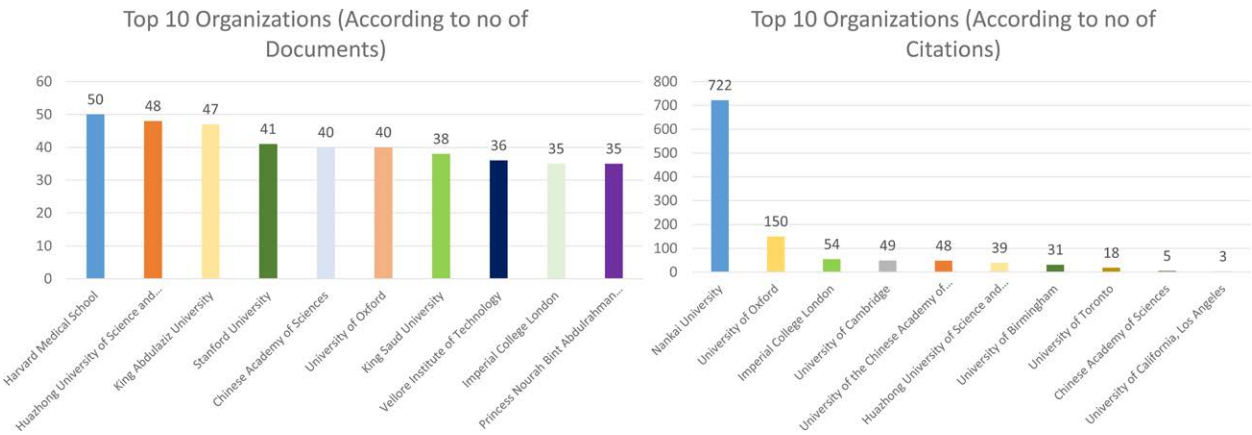


Figure 7. Network visualization map of organizations.

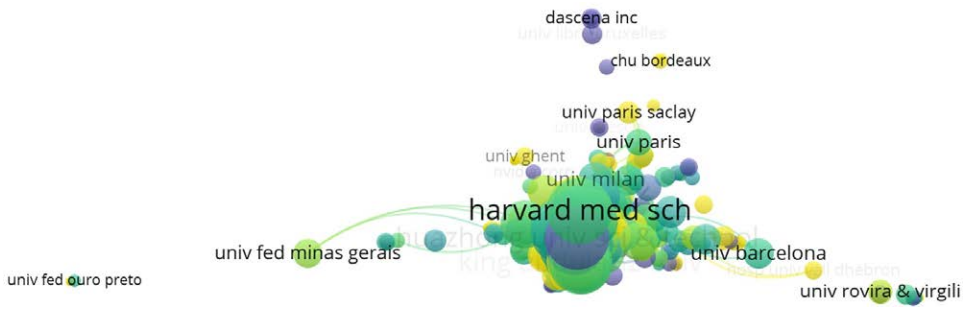


Figure 8. Top 10 organizations.

with an increase in negative emotions (anxiety, despair, and indignation) and increased sensitivity to social risks. People concerns are changing, with a stronger emphasis on health and family and less on leisure and friends. As a result of the COVID-19 outbreak, the study fills knowledge gaps about temporary changes in psychological states. The implications include giving policymakers information they may use to battle COVID-19 effectively by addressing public opinion and educating clinical staff on how to support at-risk groups and afflicted individuals.

3.6.2. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. The study investigates the effects of the control measures applied during the COVID-19 epidemic in Wuhan, China, which corresponded with the chunyun (Spring Festival) mass movement period. On January 23, 2020, China adopted substantial interventions, including widespread quarantine, travel restrictions, and case monitoring of suspected cases. In order to assess the epidemic curve, the study combines

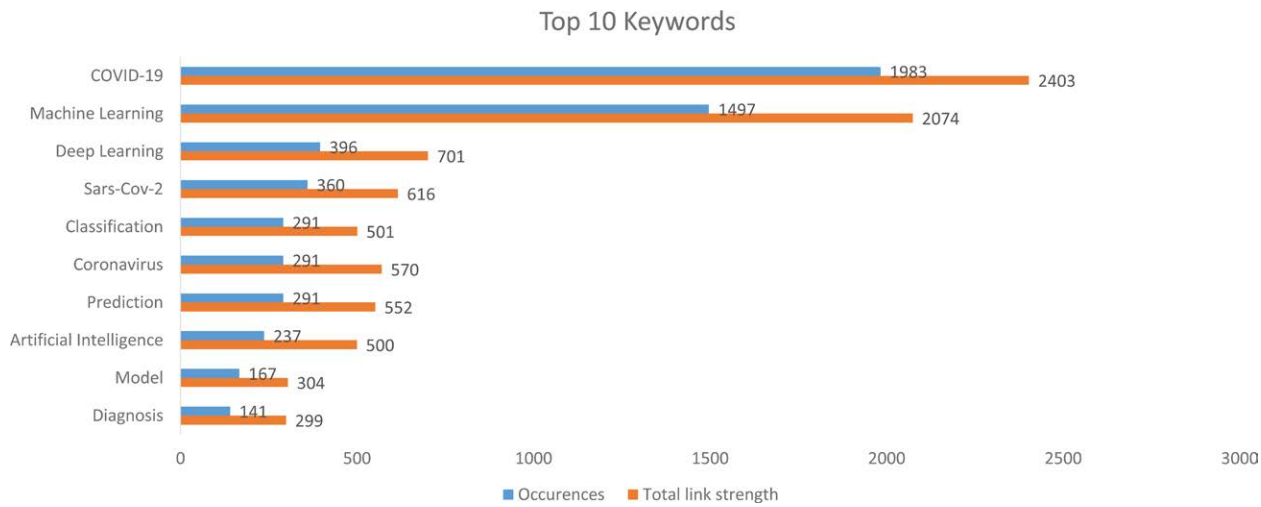


Figure 9. Top 10 keywords.

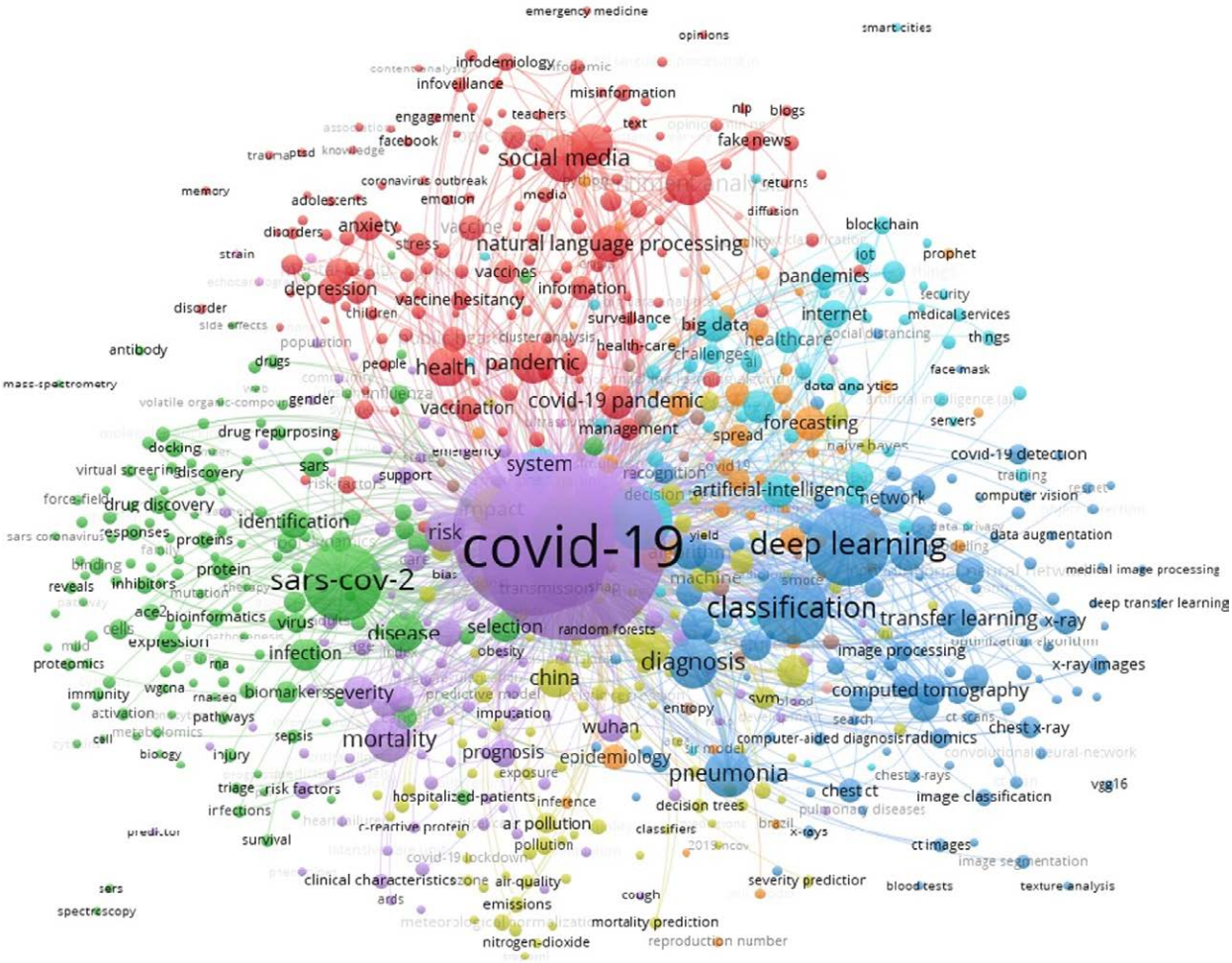


Figure 10. Network visualization map of keywords.

epidemiological data from COVID-19 and data on population migration into a Susceptible-Exposed-Infectious-Removed (SEIR) model. To forecast the course of the pandemic, AI methods trained on 2003 SARS data were applied. According to the findings, the epidemic in China should reach its height in late February and begin to subside by the end of April. Implementing control measures even 5 days later would have greatly widened the pandemic. ML forecasts show that lifting the quarantine in Hubei province would cause a second epidemic peak in mid-March and prolong the pandemic until late April. The dynamic SEIR model successfully forecasts the COVID-19 epidemic peaks and sizes, emphasizing the significance of putting control

Table 1

Top 5 most cited articles.

Article	Times cited, WoS core	Times cited, all databases	Journal	Yr
The impact of COVID-19 epidemic declaration on psychological consequences: a study on active weibo users	875	924	International Journal of Environmental Research and Public Health	2020
Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions	717	785	Journal of Thoracic Disease	2020
Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil	712	742	Science	2021
Can AI Help in Screening Viral and COVID-19 Pneumonia?	539	540	IEEE Access	2020
An interpretable mortality prediction model for COVID-19 patients	490	498	Nature Machine Intelligence	2020

AI = artificial intelligence, COVID-19 = coronavirus disease of 2019, SEIR = susceptible-exposed-infectious-removed, WoS = web of science.

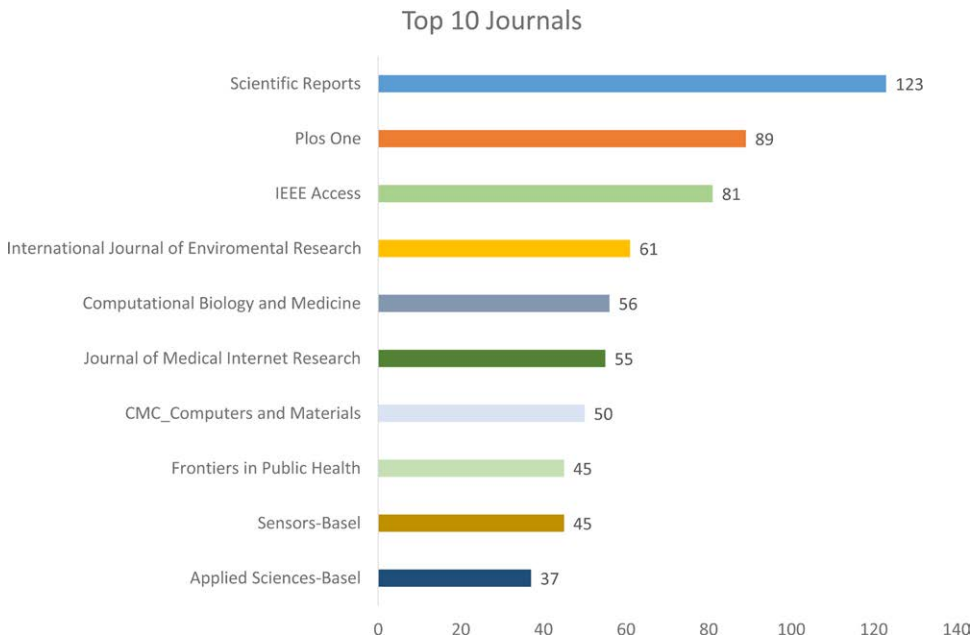


Figure 11. Top 10 journals.

measures in place on January 23, 2020, in order to minimize the extent of the epidemic in the long run.

3.6.3. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. Despite previously high infection levels, the severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2) reemerged in Manaus, Brazil, in late 2020. A novel SARS-CoV-2 variation of concern known as Lineage P.1 was discovered during the genome sequencing of viruses in Manaus from November 2020 to January 2021. Lineage P.1 had 17 mutations. Three of them, K417T, E484K, and N501Y, are changes in the spike protein that make it stick to the human ACE2 receptor better. Lineage P.1 developed around mid-November 2020, and a period of rapid molecular evolution came before it, according to a molecular clock analysis.

The study calculates that Lineage P.1 may be 1.7 to 2.4 times more transmissible than other lineages using a 2-category dynamical model that incorporates genomic and mortality data. When compared to how well it protects against non-Lineage P.1 lineages, having been infected with a non-Lineage P.1 variation protects you from Lineage P.1 infection by 54 to 79%. In order to be better prepared for pandemics, the study shows how important it is to increase global genomic surveillance of variations of concern that may be more likely to spread and avoid the immune system.

3.6.4. Can AI help in screening viral and COVID-19 pneumonia? The global pandemic of COVID-19 has caused a sizable number of deaths and illnesses. Healthcare workers need quick and reliable COVID-19 screening tools. Reverse transcription polymerase chain reaction, the main diagnostic method, has drawbacks such as high cost, low sensitivity, and the need for trained workers. In this study, AI is tested for its potential to identify COVID-19 in chest X-ray pictures. The goal is to provide a solid method employing DL algorithms that have already been taught to increase detection accuracy. A public database including 423 COVID-19, 1485 viral pneumonias, and 1579 regular chest X-ray pictures was developed. In order to train and test deep Convolutional Neural Networks (CNNs), transfer learning and picture augmentation methods were used. Two categorization schemes—normal pneumonia vs COVID-19 pneumonia and normal pneumonia versus viral pneumonia—were used to train the networks. High levels of accuracy, precision, sensitivity, and specificity were attained by the computer-aided diagnostic tool, which made it an important tool for quick and precise COVID-19 diagnosis. When there is

a pandemic and there is a greater need than supply for disease diagnosis and preventative actions, a tool like this is very helpful.

3.6.5. An interpretable mortality prediction model for COVID-19 patients. Due to the pressure on healthcare systems around the world, the study focuses on the critical need for early and precise clinical assessment of COVID-19 severity. In Wuhan, China, blood samples from 485 infected patients were used as a database to find potential disease mortality biomarkers. Three biomarkers—lactic dehydrogenase (LDH), lymphocytes, and high-sensitivity C-reactive protein (hs-CRP)—have been found by ML algorithms to be able to predict mortality in specific individuals with >90% accuracy more than 10 days in advance. Particularly high levels of LDH are emphasized as a critical marker for spotting instances necessitating rapid medical intervention. Increased LDH levels are associated with tissue breakdown observed in a number of illnesses, including lung conditions like pneumonia. In order to swiftly identify high-risk patients, facilitate prioritization, and maybe lower the mortality rate in COVID-19 instances, the study suggests a straightforward and useful decision rule.

3.7. Bibliometric analysis of top 10 journals

The top 10 journals according to the number of studies were listed in Figure 11. Scientific Reports has the most studies (123), Plos One is the second most prolific (89 studies), and IEEE Access is the third journal with 81 studies (Fig. 11).

3.7.1. Bibliometric analysis of funding agencies. The top 10 funding agencies were listed in (Fig. 12). The most effective funding agency was the United States Department of Health Human Services.

3.7.2. Bibliographic coupling.

3.7.2.1. Authors. According to the bibliographic coupling analysis, the leading 3 writers in terms of document number were Huang Tao, Feng Kaiyan, and Ashraf Imran (see Fig. 13).

According to the citation number (Fig. 13), the first 3 writers in bibliographic coupling were Xue Jia, Zhu Tingsho, and Chowdhury Muhammad E.H.

The network visualization map (Fig. 14) was produced using node weights that were determined based on the number of documents, 6 clusters, 3763 links, and a total of 94,704 link strengths.

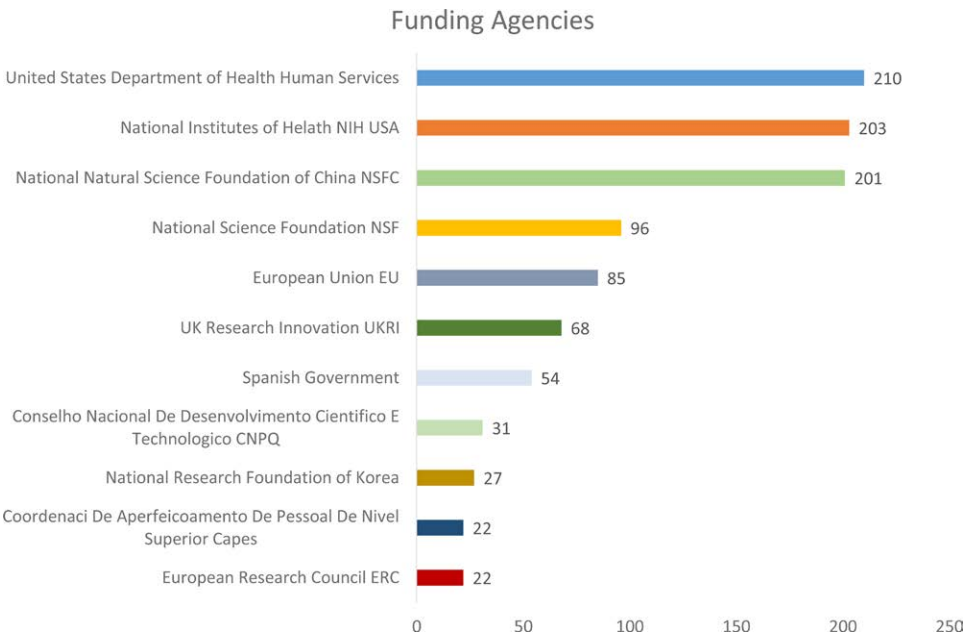


Figure 12. Top 10 funding agencies.

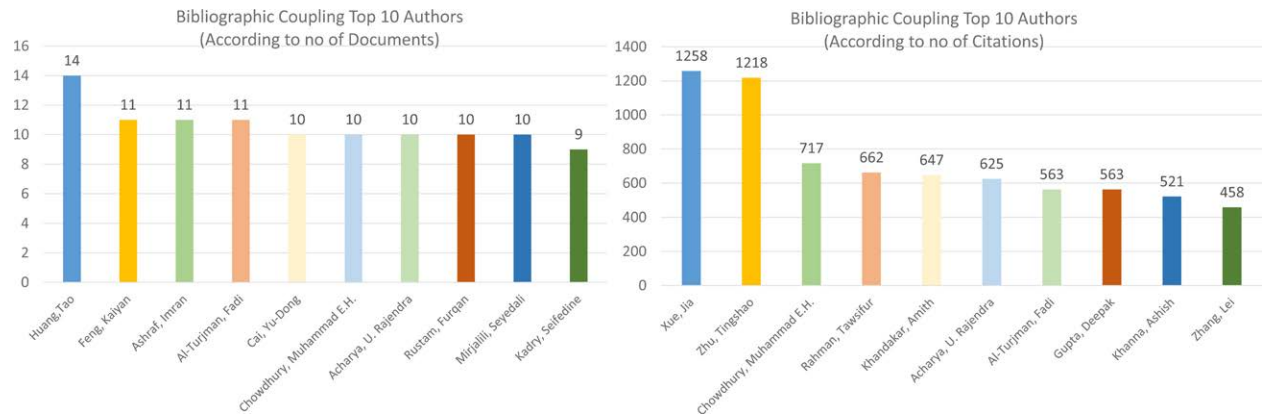


Figure 13. Bibliographic coupling top10 authors.

3.7.2.2. Documents. Chowdhurry et al (2020), Yan et al (2020) and Vaishya et al (2020) were top 3 documents according to citations (Fig. 15).[23–25]

3.7.2.3. Journals. According to Figure 16, the top 3 journals in bibliographic coupling based on document number were Scientific Reports, Plos One, and IEEE Access. On the other hand, the top 3 journals in bibliographic coupling based on citation number were IEEE Access, Science, and International Journal of Environmental Research and Public Health.

The node weights based on the number of documents, 5 clusters, 10,225 links, and 319.768 total link strengths were constructed in the network visualization map (Fig. 17).

4. Discussion

4.1. Principal results

The utilization of ML techniques is experiencing a steady rise, hence amplifying their influence on contemporary subjects, like the COVID-19 pandemic. A substantial number of research, totaling in the hundreds, have been undertaken utilizing ML techniques to investigate various aspects of the COVID-19 pandemic.[26] The objective of our work was to do a bibliometric analysis on literature related to COVID-19, employing ML techniques. The present study involved the evaluation of a total of 3559 papers that were included in the analysis. The examination of the topic involved an assessment of the published literature based on many criteria, including

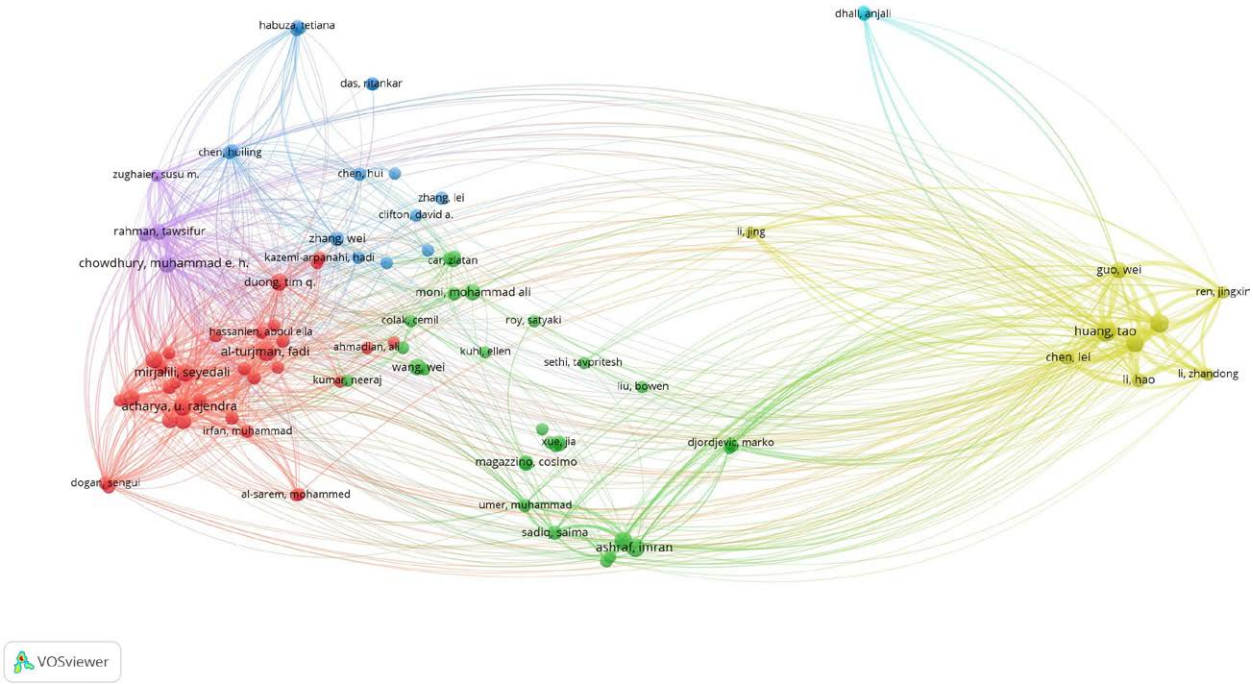


Figure 14. Network visualization map of bibliographic coupling top 10 authors.

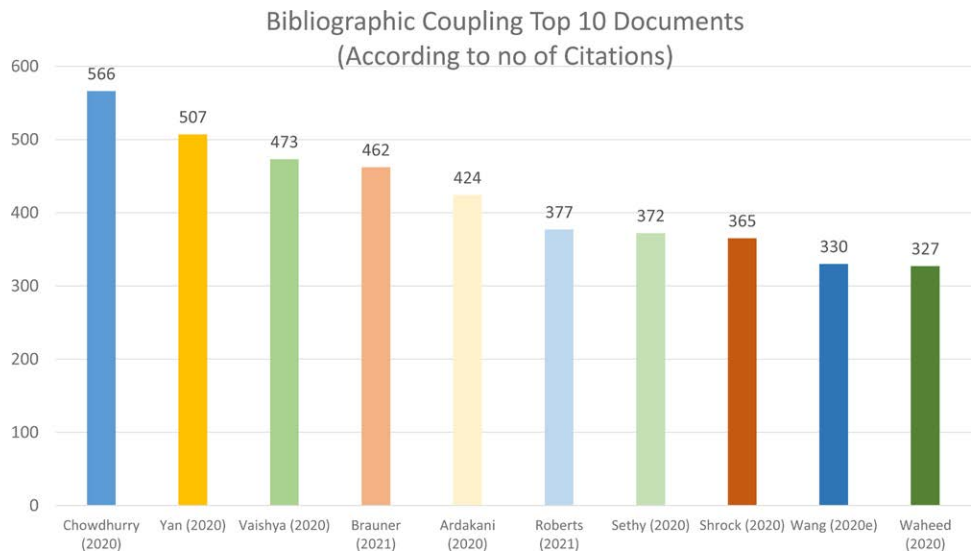


Figure 15. Bibliographic coupling top10 documents.

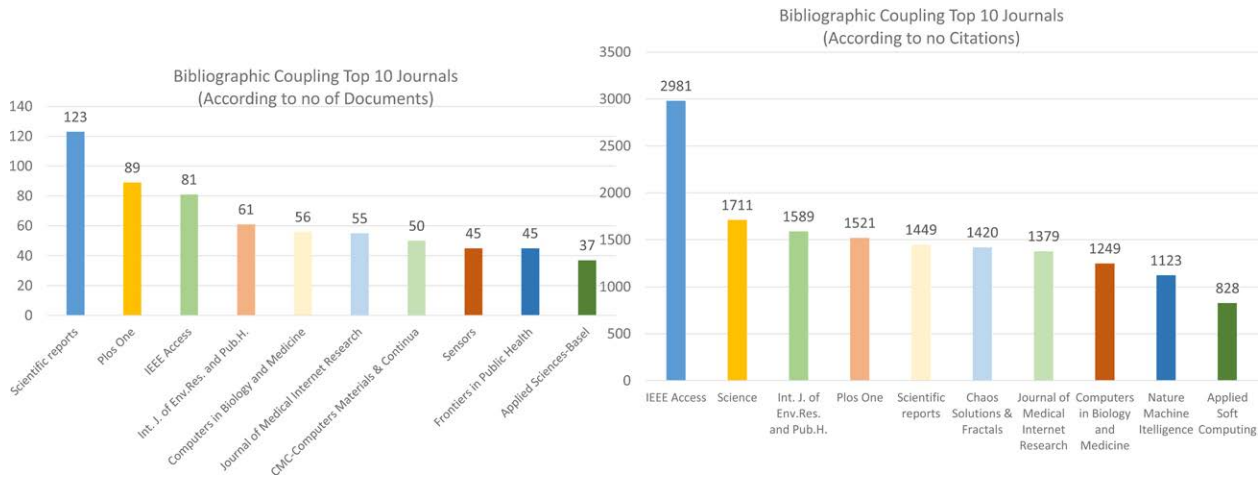


Figure 16. Bibliographic coupling top10 journals.

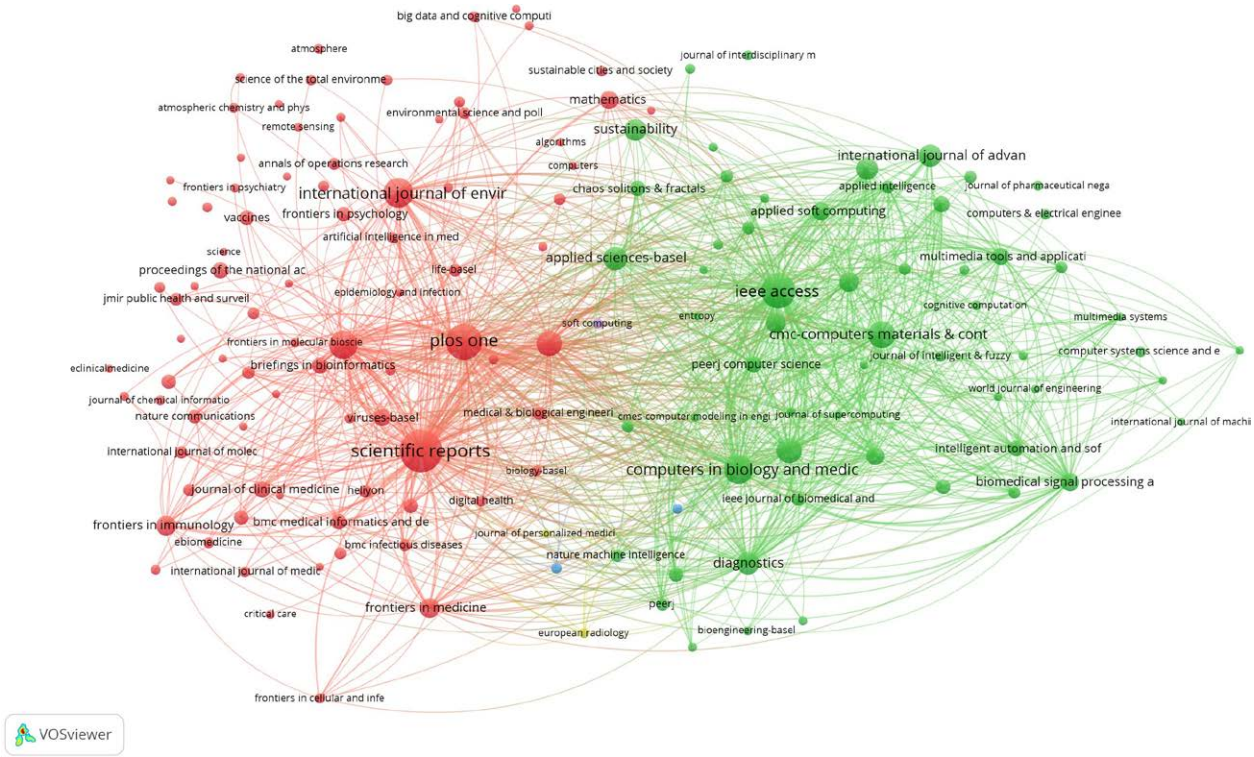


Figure 17. Network visualization map of bibliographic coupling journals.

country of origin, organizational affiliation, authorship, and keyword analysis. Based on available information, it is evident that the term "COVID-19" emerged as the prevailing keyword, aligning with expectations. Notably, the United States, China, and India emerged as the most prolific nations in terms of productivity.

Tao Huang, Fadi Al-Turjman, and Imran Ashraf (based on the number of documents) were the most prolific authors. Mazhar, Chowdhury, and Rahman were the top 3 authors by number of citations.

Harvard Medical School produced the most studies concerning ML and COVID, while Nankai University provided the most cited articles. The links between countries, organizations, and authors proved that there was a large and complicated collaboration in the context of the machine-learning techniques used in COVID-19 studies. In the current study, the top 3 most-cited studies had 875, 717, and 712 citations, respectively.

4.2. Literature comparison

Several further bibliometric analyses have been conducted on the same subject:

Steiner et al, Alavi et al, and Chiroma et al performed bibliometric analyses of COVID-19 studies using ML approaches.[27–29]

Steiner et al performed a bibliometric analysis using ML techniques applied to the coronavirus pandemic from January 2020 to June 2021. In this study, the most commonly used ML techniques were stated separately as direct COVID-19 studies, lung

XR/CT COVID-19 studies, and COVID-19 studies using social network tools.[27]

Alavi et al conducted a bibliometric analysis of the studies conducted between January 2020 and December 2022. Similar to our study, the USA, China, and India were the top 3 countries.[28] They also stated that support vector machines, computed tomography images, and transfer learning are the most commonly used ML techniques. They only included studies from PubMed, and it is suggested that researchers perform this analysis using other global databases, such as Scopus and WoS.

A study investigating the application of AI in COVID-19 was published in 2021.[30] The top 3 research areas were Computer Science Artificial Intelligence, Computer Science Information Systems, and Multidisciplinary Sciences. The first 2 were similar areas to those in our study. The top 3 countries were not changed (China, the USA, and India); however, the USA is now the leading country, while China was the first country in the study of Islam et al published in 2021.

ML, a subset of AI, has proven to have significant potential in many different sectors, including medicine and health. as data increases, ML methods can be designed to mimic human intelligence. In the field of health and medicine, a variety of different ML techniques have been applied and are being used in clinical diagnosis. Thanks to models created from datasets of medical images such as Computed Tomography (CT) Scans, Magnetic Resonance Imaging (MRI) or X-rays; prediction modules for cancer, diabetes, and other different diseases have been developed.[26] Since the beginning of the COVID-19 pandemic, governments of different countries have imposed severe sanctions to prevent people from coming together and reduce the impact of the infection. Countries have even expanded their internal restrictions and closed their borders to transit and travel to stop the spread of the virus and COVID-19. These practices have placed a significant burden on countries, adding to the disease burden on the health sector. This is why all countries and researchers are actively exploring new technologies and strategies to facilitate the tracking of this virus. AI techniques are one of the techniques that have attracted the most attention in recent years, as they can monitor, track, and assess the spread of SARS-CoV-2.[31] Identifying effective diagnostic options for the early diagnosis of COVID-19 is one of the most researched areas. For the diagnosis of SARS-CoV-2, a 2-step technique was generally used, one of which was a laboratory technique and the other a medical imaging technique. Laboratory techniques for the diagnosis of SARS-CoV-2 are known to be based on the analysis of sputum and swab samples from nasopharyngeal lesions. These samples can be evaluated in the laboratory with 2 different techniques: the first is the Rapid Antigen Test, where positivity is revealed by detecting the viral protein present in samples from infected patients. The other technique is nucleic acid amplification-based Reverse Transcriptase Polymerase Chain Reaction, which helps early detection by detecting the amplification of viral RNA at the initial stage of the disease. Radiological imaging or medical imaging techniques (CT-Scan/X-Ray), which are other methods used in diagnosis, were used because they could produce false-negative results in the laboratory. Medical imaging, namely a CT scan/X-Ray, was able to help detect the disease with 100% specificity due to specific patterns such as ground glass opacity of lung involvement. After teaching the data obtained, it is clear that if the evaluation of both laboratory and medical imaging data can be changed with the development of AI applications, faster results can be obtained.[32]

5. Conclusion

This study provides practical insights for researchers and clinicians in COVID-19 studies, including ML techniques. Using bibliometric data, researchers can obtain information regarding the most popular and productive authors and countries, the most cited manuscripts, and the most useful keywords. The data of our study showed that researchers are more oriented towards studies in this field, especially in developed countries such as the USA and China, and that the articles published here provide preliminary data for studies in this niche area and lead to more different researches. Since ML and AI modeling is a field that can develop further with new data, it should be kept in mind that both new data and new methods that will contribute to this field will continue to increase as long as the pandemic continues, but these studies will also pioneer them.

6. Limitations

- It was difficult to merge and analyze data from different databases; therefore, the WoS database was used. Scopus and PubMed can be used in future studies.
- Only English-language articles were included, which may have caused selection bias.
- The included articles were published from the database inception until March 30, 2023. However, the database is still updated, and we have missed the most recent publications.

Author contributions

Conceptualization: Alev Bakır Kayı, Mustafa Genco Erdem, Mehmet Demirci.
Data curation: Arzu Baygül Eden, Arzu Baygül Eden, Alev Bakır Kayı, Mustafa Genco Erdem, Mehmet Demirci.
Formal analysis: Arzu Baygül Eden, Alev Bakır Kayı, Mustafa Genco Erdem, Mehmet Demirci.
Methodology: Arzu Baygül Eden, Alev Bakır Kayı.
Writing – original draft: Arzu Baygül Eden, Mustafa Genco Erdem, Mehmet Demirci.
Writing – review & editing: Arzu Baygül Eden, Alev Bakır Kayı, Mustafa Genco Erdem, Mehmet Demirci.

References

[1] Amisha F, Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. J Family Med Prim Care. 2019;8:2328–31.
[2] Nichols JA, Herbert Chan HW, Baker MAB. Machine learning: applications of artificial intelligence to imaging and diagnosis. Biophys Rev. 2019;11:111–8.
[3] Sarker IH. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. SN Comput Sci. 2021;2:420.
[4] Parasher A. COVID-19: current understanding of its pathophysiology, clinical presentation and treatment. Postgrad Med J. 2021;97:312–20.
[5] Olczak J, Pavlopoulos J, Prijs J, et al. Presenting artificial intelligence, deep learning, and machine learning studies to clinicians and healthcare stakeholders: an introductory reference with a guideline and a Clinical AI Research (CAIR) checklist proposal. Acta Orthop. 2021;92:513–25.
[6] Rafique Q, Rehman A, Afghan MS, et al. Reviewing methods of deep learning for diagnosing COVID-19, its variants and synergistic medicine combinations. Comput Biol Med. 2023;163:107191.
[7] Nopour R, Shanbezadeh M, Kazemi-Arpanahi H. Predicting intubation risk among COVID-19 hospitalized patients using artificial neural networks. J Educ Health Promot. 2023;12:16.
[8] Safra M, Tamari Z, Polak P, et al. Altered somatic hypermutation patterns in COVID-19 patients classifies disease severity. Front. Immunol. 2023;14:1031914.
[9] Xue C, Xu X, Liu Z, et al. Intelligent COVID-19 screening platform based on breath analysis. J Breath Res. 2023;17:016005.
[10] Baik SM, Hong KS, Park DJ. Application and utility of boosting machine learning model based on laboratory test in the differential diagnosis of non-COVID-19 pneumonia and COVID-19. Clin Biochem. 2023;118:110584.
[11] Cropanzano R. Writing nonempirical articles for journal of management: general thoughts and suggestions. J Manage. 2009;35:1304–11.

[12] Tranfield D, Denyer D, Smart P. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. Br J Manage. 2003;14:207–22.
[13] Ellegaard O, Wallin JA. The bibliometric analysis of scholarly production: how great is the impact? Scientometrics. 2015;105:1809–31.
[14] Luukkonen T. Bibliometrics and evaluation of research performance. Ann Med. 1990;22:145–50.
[15] van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics. 2010;84:523–38.
[16] Arruda H, Silva ER, Lessa M, et al. VOSviewer and Bibliometrix. J Med Libr Assoc. 2022;110:392–5.
[17] Markscheffel B, Schröter F. Comparison of two science mapping tools based on software technical evaluation and bibliometric case studies. COLLNET J Scientometrics Inf Manage. 2021;15:365–96.
[18] Pranckute R. Web of Science (WoS) and Scopus: the titans of bibliographic information in today's academic world. Publications. 2021;9:12.
[19] Falagas ME, Pitsouni EI, Malietzis GA, et al. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. FASEB J. 2008;22:338–42.
[20] Li S, Wang Y, Xue J, et al. The impact of COVID-19 epidemic declaration on psychological consequences: a study on active weibo users. Int J Environ Res Public Health. 2020;17:2032.
[21] Yang Z, Zeng Z, Wang K, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. J Thorac Dis. 2020;12:165–74.
[22] Faria NR, Mellan TA, Whittaker C, et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. Science. 2021;372:815–21.
[23] Chowdhury MEH, Rahman T, Khandakar A, et al. Can AI help in screening viral and COVID-19 Pneumonia? IEEE Access. 2020;8:132665–76.
[24] Yan L, Zhang H-T, Goncalves J, et al. An interpretable mortality prediction model for COVID-19 patients. Nat Mach Intell. 2020;2:283–8.
[25] Vaishya R, Javaid M, Khan IH, et al. Artificial Intelligence (AI) applications for COVID-19 pandemic. Diabetes Metab Syndr. 2020;14:337–9.
[26] Mondal MRH, Bharati S, Podder P. Diagnosis of COVID-19 using machine learning and deep learning: a review. Curr Med Imaging. 2021;17:1403–18.
[27] Steiner MTA, Franco DGD, Neto PJS. Machine learning techniques applied to the coronavirus pandemic: a systematic and bibliometric analysis from January 2020 to June 2021. Rev Int Metodos Numer para Calc Diseno Ing. 2022;38:31.
[28] Alavi M, Valiollahi A, Kargari M, editors. Machine learning techniques during the COVID-19 pandemic: a bibliometric analysis. 2023 6th International Conference on Pattern Recognition and Image Analysis (IPRIA); 2023 14-16 Feb. 2023.
[29] Chiroma H, Ezugwu AE, Jauro F, et al. Early survey with bibliometric analysis on machine learning approaches in controlling COVID-19 outbreaks. PeerJ Comput Sci. 2020;6:e313.
[30] Islam MM, Poly TN, Alsinglawi B, et al. Application of artificial intelligence in COVID-19 pandemic: bibliometric analysis. Healthcare (Basel) 2021;9:441.
[31] Paul SG, Saha A, Biswas AA, et al. Combating Covid-19 using machine learning and deep learning: applications, challenges, and future perspectives. Array (N Y). 2023;17:100271.
[32] Das S, Ayus I, Gupta D. A comprehensive review of COVID-19 detection with machine learning and deep learning techniques. Health Technol (Berl). 2023;13:679–92.