Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Assessment of global research trends in the application of data science and deep and machine learning to the COVID-19 pandemic

Israel R. Orimoloye\textsuperscript{1, 2}, Olusola O. Ololade\textsuperscript{1}, Olapeju Y. Ekundayo\textsuperscript{3}, Emmanuel T. Busayo\textsuperscript{3}, Gbenga A. Afuye\textsuperscript{3}, Ahmed M. Kalumba\textsuperscript{3}, Temitope C. Ekundayo\textsuperscript{4}

\textsuperscript{1}CENTRE FOR ENVIRONMENTAL MANAGEMENT, UNIVERSITY OF THE FREE STATE, BLOEMFONTEIN, FREE STATE, SOUTH AFRICA; \textsuperscript{2}DISASTER MANAGEMENT TRAINING AND EDUCATION CENTRE FOR AFRICA, UNIVERSITY OF THE FREE STATE, BLOEMFONTEIN, FREE STATE, SOUTH AFRICA; \textsuperscript{3}DEPARTMENT OF GEOGRAPHY AND ENVIRONMENTAL SCIENCE, UNIVERSITY OF FORT HARE, ALICE, EASTERN CAPE, SOUTH AFRICA; \textsuperscript{4}DEPARTMENT OF BIOLOGICAL SCIENCES, UNIVERSITY OF MEDICAL SCIENCES, ONDO CITY, ONDO STATE, NIGERIA

1. Introduction

The coronavirus disease 2019 (COVID-19) pandemic has changed our daily lives in ways we could not have imagined since December 2019. Scientists and researchers have been employing various techniques including data science (DS), artificial intelligence, big data science, machine learning, deep learning, and other mathematic models to combat the COVID-19 pandemic. Various institutes around the world have turned their skills to help clinicians, researchers, and the medical community in their work on analyzing COVID-19 and its spread as well as the possible solution to its occurrences. Machine learning and other tools can potentially help predict risks around this pandemic. Studies have shown that machine learning research has been conducted and published on COVID-19, and early experiments are promising in terms of using this technique [1–3]. Furthermore, this can focus on how machine learning can be used in related areas and identify how it could help with risk prediction for COVID-19.

With the rapid development of innovative technology, programming, which comprises a set of instructions that produce various kinds of outputs, has been widely applied in the medical field, including organ segmentation and image enhancement and repair,
providing support for subsequent medical diagnosis [3–7]. Deep learning technologies, such as VGGNet 16, AlexNet, and ResNets with the strong ability of nonlinear modeling, have been utilized extensively in the medical field [8]. Relevant studies have been conducted on the COVID-19 pandemic and pulmonary analysis and disease prognosis globally.

Several institutions across the globe are putting efforts together to combat this pandemic. For instance, researchers at the Rensselaer Polytechnic Institute (RPI) are using machine learning to analyze the effects of social distancing, though on a more granular level [9]. Using county data from the New York State Department of Health and Mental Hygiene, RPI researchers have developed machine learning models that can predict local elements of the pandemic. These models show that the projections vary enormously from one city to another. This knowledge could relieve some of the uncertainty that is around in developing policies.

To combat and prevent this pandemic, treatment and preventive measures against this virus and gathering knowledge from scholarly and clinical research that are related to COVID-19 are crucial. A good number of clinical and observational information have been published in high-ranked outlets to assess and predict the occurrences of this pandemic. COVID-19-related published studies were retrieved from the scientific database for mapping and research trends on this pandemic using scientometric techniques. A scientometric analysis is a method of investigating published scientific information in a particular field of science through secondary analysis of the knowledge; therefore it can help researchers appreciate the earlier and existing knowledge on a specific area of interest, for example, COVID-19, and this method helps in identifying research hotspot and future research areas [10–13]. Consequently, this study aimed at evaluating global research trends that used DS, big data, machine learning, deep learning, artificial intelligence, and mathematical and statistical modeling to combat or predict COVID-19 pandemic across the globe between January and April 2020.

2. Data and methodology

2.1 Data source and identification

In order to assess the application of DS in the mitigation of COVID-19 pandemic, this study mined COVID-19-related data from the Web of Science (WoS) core collections and Scopus databases on April 18, 2020 (@23:33:22 GMT+1). First, the two databases were searched using the key term “COVID-19 OR coronavir* OR SARS-CoV-2” and setting the time span to >2019. The title field was consulted in both databases. Second, a within-result search was conducted with “model* OR machine learn* OR deep learn*”. The resultant COVID-19 dataset related to DS from WoS and Scopus was downloaded as tab-delimited (Win, UTF-8) and comma-separated formats for analysis.
2.2 Data analysis

The data was analyzed using a python programming environment based on ScientoPy package [14]. The analysis was done as described by Ruiz-Rosero et al. [14]. The analysis involved a preprocessing step in which authors’ names were normalized by removing dots, comma, and special accents from the names. The duplicated dataset from the databases was de-duplicated. In all cases of duplicated publications, the WoS version was kept while the Scopus version was removed from the study’s database. Countries’ names were also standardized for countries with at least two naming structures, such as the Republic of China: China, USA: United States; England, Scotland, and Wales: The United Kingdom; U Arab Emirates: United Arab Emirates; Russia: Russian Federation; Viet Nam: Vietnam; and Trinid & Tobago: Trinidad and Tobago.

The processed database was analyzed to answer the underlisted questions following the procedural steps outlined by Ruiz-Rosero et al. [14]:

- What are the most cited titles related to the application of DS and modeling techniques to COVID-19 pandemic?
- Where is the knowledge of DS and modeling techniques applied to COVID-19 in terms of countries and institutions?
- What are the high-frequency keywords or topics related to the application of DS and modeling techniques to COVID-19 pandemic?
- Who are the productive authors in the application of DS and modeling techniques to COVID-19 pandemic?
- What are the distributions of document types related to the application of DS and modeling techniques to COVID-19?
- What journal titles are the most productive sources related to the application of DS and modeling techniques to COVID-19?
- How are different disciplines involved in the application of DS and modeling techniques to COVID-19?
- What is the distribution of DS and modeling key terms in COVID-19 papers?

Only three performance indicators, namely, the number of papers, total citations, and h-index, were considered throughout. All analysis is inherently based on Eq. (27.1), as focal topics in a domain do have higher focal mean growth rate [14].

\[
AGR = \frac{\sum_{i=2020j_1}^{2020k_4} DS_i - DS_{i-1}}{(2020k_4 - 2020j_1)}
\]  

(27.1)

where AGR = average growth rate; 2020\(_{j_1}\) = start month; 2020\(_{k_4}\) = end month, April; and DS\(_i\) = total count of DS COVID-19-related documents in January 2020 (Table 27.1). The information in Fig. 27.1 shows the distribution of papers related to the application of DS and modeling techniques to COVID-19 in WoS and Scopus databases.
### Table 27.1 Data information.

| Document type | Total | Total citations | h-index |
|---------------|-------|-----------------|---------|
| Article       | 82    | 417             | 6       |
| Review        | 16    | 71              | 4       |
| Book chapter  | 1     | 0               | 0       |

![Figure 27.1](image)

**Figure 27.1** Distribution of papers related to the application of data science and modeling techniques to COVID-19 in Web of Science (WoS) and Scopus databases.

3. Results and discussion

Hundreds of research teams around the world are combining their efforts to collect data, analyze, and develop solutions for the COVID-19 pandemic which include identifying where and who is most at risk, diagnosing patients, developing drugs, predicting the spread of the disease, mapping where viruses come from, and understanding the virus better. The results from this study have presented various tools and techniques that have been employed in COVID-19 research between January 2020 and April 2020 by different researchers across disciplines from different countries of the world. The empirical evidence presented in this study shows that the utmost care must be taken in interpreting the result in a comparative evaluation of global research on COVID-19. A total of 82 (417 citations, 6 h-index), 16 (71 citations, 4 h-index), and 1 (no citation, no index) research articles, reviews, and book chapters, respectively, were analyzed in this study. The information in Table 27.2 presents the most cited titles related to the application of DS and modeling techniques to COVID-19 research. It was revealed that a study by Chen et al. in 2020 on “Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study” and a study by Lu et al. in 2020 on “Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding” ranked first and second with about 165 and 113 citations, respectively, and both studies were published in *Lancet*. While a study by
| Authors          | Title                                                                 | Source title                      | Cited by | Author keywords                                      | Document type | Source   | Country    |
|------------------|----------------------------------------------------------------------|----------------------------------|----------|------------------------------------------------------|---------------|----------|------------|
| Chen et al. (2020) | Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study | *Lancet*                          | 165      |                                                      | Article       | WoS      | China      |
| Lu et al. (2020)  | Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding | *Lancet*                          | 113      |                                                      | Article       | WoS      | China, Australia |
| Lai et al. (2020) | Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): the epidemic and the challenges | *International Journal of Antimicrobial Agents* | 33       | 2019-nCoV, China, COVID-19, epidemic, remdesivir, SARS-CoV-2 | Review        | Scopus   | Taiwan     |
| Wang et al. (2020) | An updated understanding of the outbreak of 2019 novel coronavirus (19-nCoV) in Wuhan, China | *Journal of Medical Virology* | 28       | Coronavirus, epidemiology, infection                 | Article       | Scopus   | China      |
| Wan et al. (2020) | Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus | *Journal of Virology*            | 26       | 2019-nCoV, SARS coronavirus, angiotensin-converting enzyme 2, animal reservoir, cross-species transmission, human-to-human transmission | Article       | WoS      | United States |
| Zhao et al. (2020) | Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak | *International Journal of Infectious Diseases* | 17       | Basic reproduction number, novel coronavirus (2019-nCoV) | Article       | WoS      | China, United States |
| Li et al. (2020)  | Coronavirus infections and immune responses                          | *Journal of Medical Virology*     | 11       | Chemokine, coronavirus, cytokines, inflammation, interferon | Review        | WoS      | China      |

*Continued*
| Authors            | Title                                                                 | Source title                     | Cited by | Author keywords                                                                 | Document type | Source | Country                  |
|--------------------|----------------------------------------------------------------------|----------------------------------|----------|---------------------------------------------------------------------------------|---------------|--------|--------------------------|
| Quilty et al.      | Effectiveness of airport screening at detecting travellers infected with the novel coronavirus (2019-nCoV) |                                  |          |                                                                                 |               |        | United Kingdom           |
| Benvenuto et al.   | The 2019-new coronavirus epidemic: evidence for virus evolution      | *Journal of Virology*            | 7        | Coronavirus, epidemiology, macromolecular design, SARS coronavirus              | Article       | WoS     | Italy, Brazil            |
| Zhao et al.        | Estimating the unreported number of novel coronavirus (2019-nCoV) cases in China in the first half of January 2020: A data-driven modelling analysis of the early outbreak | *Journal of Clinical Medicine*    | 6        | Novel coronavirus, outbreak, modeling, underreporting, reproduction number, China | Article       | WoS     | China, United States     |
| Gilbert et al.     | Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study | *Lancet*                         | 6        |                                                                                 | Article       | WoS     | Belgium, France, United States, Cote d'Ivoire, United Kingdom |
Lai et al. in 2020 on “Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): the epidemic and the challenges” and a study by Wang et al. in 2020 on “Updated understanding of the outbreak of 2019 novel coronavirus (2019-nCoV) in Wuhan, China” ranked third and fourth received 33 and 28 citations, respectively, during the survey period, and both studies were published as articles in the International Journal of Antimicrobial Agents. It was noted that these top titles used various related author keywords, which include 2019-nCoV, China, COVID-19, epidemic, remdesivir, SARS-CoV-2, coronavirus, epidemiology, infection 2019-nCoV, SARS coronavirus, angiotensin-converting enzyme 2, animal reservoir, cross-species transmission, and human-to-human transmission in COVID-19 studies between January and April 2020.

More so, country and institutional participation in the application of DS and modeling techniques to COVID research is presented in Table 27.3. The results from the study reveal that China, the United States, the United Kingdom, Australia, and Canada ranked first, second, third, fourth, and fifth with about 39 (366 citations),

| Position | Country         | Total citations | h-index | Institution with country            | Total citations | h-index |
|----------|-----------------|-----------------|---------|-------------------------------------|-----------------|---------|
| 1        | China           | 39              | 6       | Chinese Acad Sci, China             | 4               | 278     | 2       |
| 2        | United States   | 32              | 5       | Hong Kong Polytech Univ, China      | 3               | 23      | 2       |
| 3        | United Kingdom  | 13              | 4       | Univ Hong Kong, China               | 3               | 23      | 2       |
| 4        | Australia       | 7               | 2       | Univ Oxford, United Kingdom         | 3               | 12      | 2       |
| 5        | Canada          | 7               | 1       | Wuhan Univ, China                   | 3               | 11      | 1       |
| 6        | Italy           | 6               | 2       | Capital Med Univ, China             | 2               | 0       | 0       |
| 7        | India           | 5               | 1       | Chinese Univ Hong Kong, China       | 2               | 23      | 2       |
| 8        | South Korea     | 5               | 0       | Georgia State Univ, United States   | 2               | 5       | 1       |
| 9        | Germany         | 4               | 1       | Harbin Engn Univ, China             | 2               | 0       | 0       |
| 10       | Belgium         | 3               | 6       | Hokkaido Univ, Japan                | 2               | 4       | 1       |
| 11       | Japan           | 3               | 9       | Huaiyin Normal Univ, China          | 2               | 23      | 2       |
| 12       | France          | 2               | 7       | Japan Sci & Technol Agcy, Japan     | 2               | 4       | 1       |
| 13       | Iran            | 2               | 0       | Peking Univ, China                  | 2               | 0       | 0       |
| 14       | Netherlands     | 2               | 1       | Shanghai Jiao Tong Univ, China      | 2               | 167     | 2       |
| 15       | Saudi Arabia    | 2               | 8       | Shanghai Normal Univ, China         | 2               | 23      | 2       |
| 16       | Sweden          | 2               | 2       | South China Agr Univ, China         | 2               | 0       | 0       |
| 17       | Switzerland     | 2               | 1       | Southern Med Univ, China            | 2               | 11      | 1       |
| 18       | Taiwan          | 2               | 33      | Univ Calif Los Angeles, United States| 2               | 116     | 1       |
| 19       | Turkey          | 2               | 0       | Univ Michigan, United States        | 2               | 23      | 2       |
| 20       |                 |                 |         | Uni Sydney, Australia               | 2               | 116     | 2       |
32 (79 citations), 13 (33 citations), 7 (116 citations), and 7 (3 citations) articles, respectively. While Italy, India, South Korea, Germany, and Belgium with about 6 (10 citations), 5 (3 citations), 5 (0 citations), 4 (4 citations), and 3 (6 citations) articles stand at sixth, seventh, eighth, ninth, and tenth ranks, respectively.

Instructional relevance information in Table 27.3 reveals that Chinese Acad Sci (4 articles, 278 citations) in China; Hong Kong Polytech Univ in China (3 articles, 23 citations); Univ Hong Kong (3 articles, 23 citations) in China; Univ Oxford United Kingdom (3 articles, 12 citations); Wuhan Univ, China (3 articles, 11 citations); Capital Med Univ, China (2 articles, 0 citation); Chinese Univ Hong Kong, China (2 articles, 23 citations); Georgia State Univ, United States (2 articles, 5 citations); Harbin Engn Univ, China (2 articles, 0 citations); and Hokkaido Univ, Japan (2 articles, 4 citations) ranked first, second, third, fourth, fifth, sixth, seventh, eighth, ninth, and tenth, respectively. This study also shows that the top productive nations and institutions are the countries that are most affected by COVID-19 pandemic between January and April 2020.

High-frequency author keywords and keyword plus related to the application of DS and modeling techniques on COVID-19-related research during the period of study are presented in Table 27.4. It was noted that “COVID-19” and “coronavirus” ranked first

| Position | Author keywords | Total | h-index | Both keywords | Total | h-index |
|----------|----------------|-------|---------|---------------|-------|---------|
| 1        | COVID-19       | 36    | 3       | COVID-19      | 41    | 4       |
| 2        | Coronavirus    | 21    | 4       | Coronavirus   | 26    | 4       |
| 3        | SARS-CoV-2     | 18    | 3       | Human         | 22    | 3       |
| 4        | 2019-nCoV      | 7     | 3       | China         | 18    | 4       |
| 5        | Novel coronavirus | 7  | 3       | SARS-CoV-2    | 18    | 3       |
| 6        | China          | 6     | 3       | EPIDEMIC      | 17    | 4       |
| 7        | Epidemiology   | 6     | 3       | Article       | 15    | 3       |
| 8        | Basic reproduction number | 5 | 2       | Coronavirus infection | 15 | 4 |
| 9        | Epidemic       | 5     | 2       | Coronavirus disease 2019 | 14 | 3 |
| 10       | Machine learning | 4 | 1       | Nonhuman      | 13    | 4       |
| 11       | Pneumonia      | 4     | 2       | Severe acute respiratory syndrome Coronavirus 2 | 13 | 3 |
| 12       | Coronavirus disease 2019 (COVID-19) | 4 | 1       | Coronavirinae | 11 | 4 |
| 13       | Travel         | 4     | 2       | SARS          | 11    | 3       |
| 14       | Artificial intelligence | 3 | 0       | Virus pneumonia | 11 | 4 |
| 15       | Deep learning  | 3     | 0       | Betacoronavirus | 10 | 3 |
| 16       | Outbreak       | 3     | 2       | Coronavirus infections | 10 | 4 |
| 17       | Remdesivir     | 3     | 1       | Humans        | 10    | 4       |
| 18       | SARS           | 3     | 2       | Pneumonia, viral | 10 | 4 |
| 19       | Wuhan          | 3     | 2       | SARS coronavirus | 10 | 4 |
| 20       | Migration      | 3     | 1       | Priority journal | 10 | 3 |
with 36 (3 h-index) and 21 (4 h-index) appearances in both author keywords and keyword plus in COVID-19-related research. Studies have shown that keywords represent the author’s opinion of the most important terms or words in any given research article. More so, keyword and keyword plus analysis can potentially detect trending research topics both currently and in the past and the likely future research direction [13,15,16]. More so, keywords analysis can answer several interesting questions and research key themes in research articles that can identify research evolution and hotspot. The analysis of top productive authors in the application of DS and modeling techniques to COVID-19 research has revealed that Yang L. with 5 articles (23 citations, 2 h-index), Wang Y. with 4 articles (9 citations, 2 h-index), He D.H. with about 3 articles (23 citations, 2 h-index), Li S.Y. with 3 articles (1 citation, 1 h-index), and Li S. with 3 articles (0 citations, 0 h-index) ranked first, second, third, fourth, and fifth, respectively (Table 27.5).

Table 27.6 presents the most influential and relevant sources of research on COVID-19 that have applied various techniques, which include DS, big data, machine learning, deep

| Table 27.5  | Productive authors in the application of data science and modeling techniques to COVID-19. |
|-------------|------------------------------------------------------------------------------------------|
| Author      | Total | Citations | h-index |
| Yang L.     | 5     | 23        | 2       |
| Wang Y.     | 4     | 9         | 2       |
| He D.H.     | 3     | 23        | 2       |
| Li S.       | 3     | 0         | 0       |
| Li S.Y.     | 3     | 1         | 1       |

| Table 27.6  | Most productive journal sources related to the application of data science and modeling techniques to COVID-19. |
|-------------|-----------------------------------------------------------------------------------------------------------|
| Position    | Source title                                                                                       | Total | Total citations | h-index |
| 1           | Journal of Clinical Medicine                                                                       | 6     | 13              | 2       |
| 2           | Journal of Medical Virology                                                                       | 5     | 50              | 4       |
| 3           | Eurosurveillance                                                                                  | 4     | 13              | 2       |
| 4           | Mathematical Biosciences and Engineering                                                            | 4     | 0               | 0       |
| 5           | Lancet                                                                                           | 3     | 284             | 3       |
| 6           | Cell Discovery                                                                                     | 2     | 2               | 1       |
| 7           | Emerging Microbes & Infections                                                                   | 2     | 0               | 0       |
| 8           | Infectious Diseases of Poverty                                                                    | 2     | 0               | 0       |
| 9           | International Journal of Antimicrobial Agents                                                     | 2     | 33              | 1       |
| 10          | Journal of Infection in Developing Countries                                                       | 2     | 3               | 1       |
| 11          | Journal of Virology                                                                               | 2     | 27              | 1       |
| 12          | J. of Shanghai Jiaotong University (Science)                                                       | 2     | 0               | 0       |
| 13          | Medical Science Monitor                                                                            | 2     | 0               | 0       |
| 14          | One Health                                                                                       | 2     | 2               | 1       |
| 15          | Science of the Total Environment                                                                  | 2     | 0               | 0       |
learning, artificial intelligence, and mathematic and statistical modeling between January and April 2020. The result from the study reveals the top journals or sources with the most published research articles on COVID-19-related studies. These relevant sources cover a range of subjects in their respective articles. From Table 27.6, it was derived that quite a diversity with respect to the scholarly fields of health and mathematic biosciences, as well as engineering and surveillance. The diversity of outlets is also evident in the results for disciplines, as can be seen in Table 27.7. Interestingly, the open-access journal, *Journal of Clinical Medicine*, with 6 articles (13 citations) ranked the first most influential journal in COVID-19 research. This journal has been publishing since 2012 and particularly publishes within science and medicine. The top 15 journals are presented in Table 27.6: second position, *Journal of Medical Virology*; third, *Eurosurveillance*; fourth, *Mathematical Biosciences and Engineering*; and fifth, *Lancet* with 3 articles has the highest number of citations of about 284 during the survey period. Other relevant sources on COVID research and their respective statistics are presented in Table 27.6. Discoveries from this study reveal COVID-19 research is more preliminary. Achieving significant insights will require a mix of domain expertise from multiple fields, and there is already a push for better international collaboration and tracking of COVID-19 [17]. For instance, the use of big data, machine learning, deep learning, artificial intelligence, and mathematic and statistical modeling might yield a superficially practical solution but could be ineffective without the involvement of (international) medical and biotechnology expert interpretations. This would also have implications for emerging innovations (as it is unlikely that healthcare practitioners would interact with technologies built without medical expertise). So it is necessary to quickly put together cohorts with complementary expertise. This also brings many challenges, such as ensuring a team is consistent in interpreting things such as ethics, benefits, and risks.

3.1 Distribution of subject areas

The distribution of the top nine disciplines in the application of DS and modeling techniques to COVID-19 studies is listed in Table 27.7: General and internal medicine is

| Subject                                  | Total | Total citations | h-index |
|------------------------------------------|-------|----------------|---------|
| General & internal medicine              | 11    | 297            | 4       |
| Infectious diseases                      | 9     | 28             | 3       |
| Virology                                 | 6     | 45             | 3       |
| Microbiology                             | 5     | 0              | 0       |
| Mathematical & computational biology     | 4     | 0              | 0       |
| Research & experimental medicine         | 3     | 0              | 0       |
| Computer Science                         | 2     | 0              | 0       |
| Healthcare sciences & services           | 2     | 0              | 0       |
| Immunology                               | 2     | 0              | 0       |
ranked number 1 with 11 articles, 297 total citations, and an h-index of 4, followed by infectious diseases, with 9 articles, 28 total citations, and an h-index of 3. Academic outputs using DS and modeling techniques to COVID-19 focused majorly on health science subject areas, especially the areas concerned with restoring and maintaining human health through the treatment of disease and injury. These disciplines fundamentally address how knowledge is applied to the prevention and cure of diseases in the body systems. The contribution of the top subjects that applied DS and modeling to COVID-19 was expanded in Fig. 27.2. The analysis shows demography, which is the science of populations as the only subject area focusing on social science; this implies that researchers were focusing on DS application to COVID-19 more from the perspectives of health sciences and natural sciences. Other domain-indicated subjects including public, environmental, and occupational health; virology; research and experimental medicine; and life sciences also published numerous papers using DS because of their sensitivity in these fields to COVID-19, and its major impacts thereof mainly happened in these subject areas. It is essential at this time for health professional researchers to work with social scientists to inform effective policies [17].

The synergy of various research areas would help in addressing salient issues on the outbreak of COVID-19 and perhaps to understand future events of this manner. There is

FIGURE 27.2 Top 30 productive subject areas on the application of data science and modeling techniques to COVID-19.
a shortage of studies on subject areas related to human behavioral space, and this should be very crucial in understanding person-to-person transmission of this virus, hence researchers within subject areas related to human behavioral space should do more in identifying and understanding any possible barriers in flattening the curve of the pandemic [18–20]. Future research should address potential variation in social issues very critically, hence the ripple effects of the virus outbreak are basically on the social life of the populace, which has led to the popular preventive measure of COVID-19 tagged “social distancing” i.e., avoiding social meetings, self-isolating, etc. [21]. The outbreak of the virus and its future implications require special attention, which should provide a paradigm shift in different works of life, especially human movements across various territories [22].

### 3.2 Research key terms

Keywords or key terms reflect the authors’ aims and objectives and summarize the key intentions and interests of a paper; therefore the distribution of DS and modeling key terms in COVID-19-related papers is key to investigate hot topics and research erudition related to this subject area (Fig. 27.3). The information in Fig. 27.3 reveals the number and percentage, as well as the distribution, of DS and modeling techniques to COVID-19 before 2020 and during 2020 (January to April 2020). In Table 27.8, the top nine author keywords, both keywords, total articles, total citations, and h-index were summarized. “Model” has the highest frequency with 19 papers, 11 total citations, and an h-index of 2; on the other hand, it appeared number one on both keywords with 37 papers, 47 citations,
and an h-index of 4. Bibliometrics generally applies Price’s law to estimate authors’ influence and output in a particular field of study, which can also determine the highest and lowest occurrence of important key terms [23].

### 3.2.1 Model

Table 27.8 shows that “Model” is a taxonomy of keywords with the highest occurrence; this implies that modeling has become increasingly significant in DS application to COVID-19 studies. For example, a “stochastic transmission model,” adopted to COVID-19, was used to quantify the possible efficiency of tracing contact and isolation of infected people to control a severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)-like pathogen [24]. In another instance, validated phenomenological models that have been applied in past outbreaks were used to generate and assess short-term forecasts of the cumulative number of confirmed reported cases in the Hubei province, the epicenter of the pandemic, and for the overall cases in China, not including the province of Hubei, this model was hinged on three other models, which are generalized logistic model, the Richards growth model, and a subepidemic wave model [25].

Modeling is the act of creating physical, philosophical, and conceptual structures and the relationship between data elements [26]. This creates an avenue to establish goals and provide information for end-users. Modeling in the case of COVID-19 requires sufficient historical data. However, we should be cognizant that history often repeats itself just the same way it has happened in the past. Hence all hands must be on deck in modeling future events of similar pandemic as it relates to COVID-19. Answering the questions related to the global impact of the novel coronavirus (COVID-19) requires accurate analysis as well as modeling recorded cases, deaths, and number of recoveries [27,28].

### 3.2.2 Machine learning

This computer program focuses on accessing data and automating processes based on inputs without explicit programming. The efficacy of machine learning cannot be

| Position | Author keywords       | Total citations | h-index | Both keywords       | Total citations | h-index |
|----------|-----------------------|-----------------|---------|---------------------|-----------------|---------|
| 1        | Model                 | 19              | 11      | 2                   | 1   | Model | 37  | 47 | 4   |
| 2        | Machine learning      | 4               | 1       | 1                   | 2   | Machine learning | 5   | 1  | 1  |
| 3        | Mathematical          | 4               | 1       | 1                   | 3   | Artificial intelligence | 4   | 0  | 0  |
| 4        | Artificial intelligence | 3   | 1       | 0                   | 4   | Mathematical | 4   | 1  | 1  |
| 5        | Deep learning         | 3               | 0       | 0                   | 5   | Deep learning | 3   | 0  | 0  |
| 6        | Regression            | 1               | 0       | 0                   | 6   | Regression | 3   | 0  | 0  |
| 7        | Active learning       | 1               | 0       | 0                   | 7   | Computational | 1   | 0  | 0  |
| 8        | Computational         | 0               | 0       | 0                   | 8   | Active learning | 1   | 0  | 0  |
| 9        | Statistical model     | 0               | 0       | 0                   | 9   | Statistical model | 0   | 0  | 0  |

Table 27.8 Distribution of data science and modeling key terms in COVID-19 papers.
overemphasized in DS application to COVID-19 based on its ability to access data and apply it accordingly. Rao and Vazquez [29] proposed to use machine learning algorithms to improve possible case identifications of COVID-19 quicker with the use of a mobile phone-based web survey in order to flatten the curve of the virus in vulnerable populations.

3.2.3 Artificial intelligence
The simulation of human intelligence in programmed machines is the thrust of artificial intelligence. This is also applied to machines that solve problems and show traits associated with human actions. Artificial intelligence with the use of deep learning technology has shown great success, with the potential to accurately detect COVID-19 and distinguish it from communal acquired pneumonia and other lung infections [30].

3.2.4 Other key terms
According to Table 27.8, the other hot keywords in the application of DS to COVID-19 also include mathematical, deep learning, regression, active learning, computational and statistical models, etc. In general, mathematic and computational statistical models are one of the approaches that are vital in data analytics to navigate the nebulous nature of big data. Hence, this area has become a forte of many researchers. Regional limitation of researchers is not a problem with the use of DS in COVID-19 studies; so far the required data is available, and studies can be conducted anywhere across the world.

4. Conclusion
This study shows the significant focus of researchers on the application of DS to address COVID-19 outbreak, and studies recorded in the analysis majorly focused on health science subject areas. Country and institutional participation in the application of DS and modeling techniques to COVID research was identified and presented in the study. The results from the study reveal that China, the United States, the United Kingdom, Australia, and Canada ranked first, second, third, fourth, and fifth, respectively. While Italy, India, South Korea, Germany, and Belgium stand at sixth, seventh, eighth, ninth, and tenth, respectively. This ranking was based on the number of published articles, the number of citations, and h-index during the survey period.

The key terms show the classification of the major approach used in addressing COVID-19 in the area of DS application, which was found mainly in nine categories such as model, machine learning, artificial intelligence, mathematic, deep learning, regression, computational, active learning, and statistical model. “Model” is the major keyword found in research published in the year 2020 applying DS to COVID-19; it appears to be the most promising approach for developing physical, theoretical, or computer-based simplifications to address COVID-19.
References

[1] A. Alimadadi, S. Aryal, I. Manandhar, P.B. Munroe, B. Joe, X. Cheng, Artificial intelligence and machine learning to fight COVID-19, Physiol. Genom. (2020).

[2] M. Barstugan, U. Ozkaya, S. Ozturk, Coronavirus (Covid-19) Classification Using CT Images by Machine Learning Methods, arXiv, 2020 preprint arXiv:2003.09424.

[3] C. Butt, J. Gill, D. Chun, B.A. Babu, Deep learning system to screen coronavirus disease 2019 pneumonia, Appl. Intell. (2020) 1.

[4] M. Farooq, A. Hafeez, Covid-Resnet: A Deep Learning Framework for Screening of Covid19 From Radiographs, arXiv, 2020 preprint arXiv:2003.14395.

[5] H.S. Maghdid, K.Z. Ghafoor, A.S. Sadiq, K. Curran, K. Rabie, A Novel AI-Enabled Framework to Diagnose Coronavirus Covid 19 Using Smartphone Embedded Sensors: Design Study, arXiv, 2020 preprint arXiv:2003.07434.

[6] A. Narin, C. Kaya, Z. Pamuk, Automatic Detection of Coronavirus Disease (Covid-19) Using X-Ray Images and Deep Convolutional Neural Networks, arXiv, 2020 preprint arXiv:2003.10849.

[7] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, D. Shen, Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for Covid-19, IEEE Rev. Biomed. Eng. (2020).

[8] H. Jiang, F. Gao, X. Duan, Z. Bai, Z. Wang, X. Ma, Y.W. Chen, Transfer learning and fusion model for classification of epileptic PET images, in: Innovation in Medicine and Healthcare Systems, and Multimedia, Springer, Singapore, 2019, pp. 71–79.

[9] HIA, Health IT Analytics, Xtelligent Healthcare Media, 2020.

[10] W. Li, Y. Zhao, Bibliometric analysis of global environmental assessment research in a 20-year period, Environ. Impact Assess. Rev. 50 (2015) 158–166.

[11] G. Mao, H. Zou, G. Chen, H. Du, J. Zuo, Past, current and future of biomass energy research: a bibliometric analysis, Renew. Sustain. Energy Rev. 52 (2015) 1823–1833.

[12] T.C. Ekundayo, A.I. Okoh, A global bibliometric analysis of Plesiomonas-related research (1990–2017), PLoS One 13 (11) (2018).

[13] I.R. Orimoloye, O.O. Ololade, Global trends assessment of environmental health degradation studies from 1990 to 2018, Environ. Dev. Sustain. (2020) 1–4.

[14] J. Ruiz-Rosero, G. Ramirez-Gonzalez, J. Viveros-Delgado, Software survey: ScientoPy, a scientometric tool for topics trend analysis in scientific publications, Scientometrics 121 (2) (2019) 1165–1188.

[15] B. Pesta, J. Fuerst, E.O. Kirkegaard, Bibliometric keyword analysis across seventeen years (2000–2016) of intelligence articles, J. Intell. 6 (4) (2018) 46.

[16] I.R. Orimoloye, O.O. Ololade, Potential implications of gold-mining activities on some environmental components: a global assessment (1990 to 2018), J. King Saud Univ. Sci. (2020).

[17] C.A. Harper, L.P. Satchell, D. Fido, R.D. Latzman, Functional fear predicts public health compliance in the COVID-19 pandemic, Int. J. Ment. Health Addiction (2020) 1.

[18] J.A. Everett, C. Colombatto, V. Chituc, W.J. Brady, M. Crockett, The effectiveness of moral messages on public health behavioral intentions during the COVID-19 pandemic, PsyArXiv Preprints (2020).

[19] E.A. Holmes, R.C. O'Connor, V.H. Perry, I. Tracey, S. Wessely, L. Arseneault, C. Ballard, H. Christensen, R.C. Silver, I. Everall, T. Ford, Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science, Lancet Psychiatr. (2020).

[20] K. Prem, Y. Liu, T.W. Russell, A.J. Kucharski, R.M. Eggo, N. Davies, S. Flasche, S. Clifford, C.A. Pearson, J.D. Munday, S. Abbott, The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study, Lancet Public Health (2020).
[21] M. Greenstone, V. Nigam, Does Social Distancing Matter?, University of Chicago, Becker Friedman Institute for Economics Working Paper, 2020. Mar 25 (2020-26).

[22] K.C. Santosh, AI-driven tools for coronavirus outbreak: need of active learning and cross-population train/test models on multitudinal/multimodal data, J. Med. Syst. 44 (5) (2020) 1–5.

[23] D.J. de Solla Price, D. Beaver, Collaboration in an invisible college, Am. Psychol. 21 (11) (1966) 1011.

[24] J. Hellewell, S. Abbott, A. Gimma, N.I. Bosse, C.I. Jarvis, T.W. Russell, J.D. Munday, A.J. Kucharski, W. J. Edmunds, F. Sun, S. Flasche, Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts, Lancet Glob. Health (2020).

[25] K. Roosa, Y. Lee, R. Luo, A. Kirpich, R. Rothenberg, J.M. Hyman, P. Yan, G. Chowell, Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020, Infect. Dis. Model. 5 (2020) 256–263.

[26] B. Succar, M. Kassem, Macro-BIM adoption: conceptual structures, Autom. ConStruct. 57 (2015) 64–79.

[27] A. Wilder-Smith, C.J. Chiew, V.J. Lee, Can we contain the COVID-19 outbreak with the same measures as for SARS? Lancet Infect. Dis. (2020).

[28] C.S. Currie, J.W. Fowler, K. Kotiadis, T. Monks, B.S. Onggo, D.A. Robertson, A.A. Tako, How simulation modelling can help reduce the impact of COVID-19, J. Simulat. (2020) 1–5.

[29] A.S. Rao, J.A. Vazquez, Identification of COVID-19 can be quicker through artificial intelligence framework using a mobile phone–based survey when cities and towns are under quarantine, Infect. Contr. Hosp. Epidemiol. (2020) 1–5.

[30] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song, K. Cao, Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT, Radiology (2020). https://doi.org/10.1148/radiol.2020200905.