Early survey with bibliometric analysis on machine learning approaches in controlling coronavirus

Haruna Chiroma [1] , **Absalom E. Ezugwu** [Corresp., 2] , **Fatsuma Jauro** [3] , **Mohammed A Al-Garadi** [4] , **Idris N Abdullahi** [5] , **Liyana Shuib** [6]

[1] Future Technology Research Center, National Yunlin University of Science and Technology, Yulin, Taiwan

[2] School of Computer Science, University of KwaZulu-Natal, Durban, KwaZulu-Natal, South Africa

[3] Department of Computer Science, Faculty of Science, Ahmadu Bello University, Zaria, Nigeria

[4] Department of Biomedical Informatics, Emory University, Atlanta, United States

[5] Department of Medical Laboratory Science, College of Medical Sciences, Ahmadu Bello University, Zaria, Nigeria

[6] Department of Information System, Universiti Malaya, Kuala Lumpur, Malaysia

Corresponding Author: Absalom E. Ezugwu
Email address: ezugwua@ukzn.ac.za

Background and Objective: The COVID-19 pandemic has caused severe mortality across the globe with the USA as the current epicenter of the COVID-19 epidemic even though the initial outbreak was in Wuhan, China. Many studies successfully applied machine learning to fight COVID-19 pandemic from a different perspective. To the best of the authors' knowledge, no comprehensive survey with bibliometric analysis has been conducted yet on the adoption of machine learning to fight COVID-19. Therefore, the main goal of this study is to bridge this gap by carrying out an in-depth survey with bibliometric analysis on the adoption of machine learning-based technologies to fight COVID-19 pandemic from a different perspective, including an extensive systematic literature review and bibliometric analysis. **Methods:** We applied a literature survey methodology to retrieved data from academic databases and subsequently employed a bibliometric technique to analyze the accessed records. Besides, the concise summary, sources of COVID-19 datasets, taxonomy, synthesis and analysis are presented in this study. It is found that the Convolutional Neural Network (CNN) is mainly utilized in developing COVID-19 diagnosis and prognosis tools, mostly from chest X-ray and chest CT scan images. Similarly, in this study, we performed a bibliometric analysis of machine learning-based COVID-19 related publications in the Scopus and Web of Science citation indexes. Finally, we propose a new perspective for solving the challenges identified as direction for future research. We believe the survey with bibliometric analysis can help researchers easily detect areas that require further development and identify potential collaborators. **Results:** The findings of the analysis presented in this paper reveal that machine learning-based COVID-19 diagnose tools received the most considerable attention from researchers. Specifically, the

analyses of results show that energy and resources are more dispenses towards COVID-19 automated diagnose tools while COVID-19 drugs and vaccine development remains grossly underexploited. Besides, the machine learning-based algorithm that is predominantly utilized by researchers in developing the diagnostic tool is CNN mainly from X-rays and CT scan images. **Conclusions:** The challenges hindering practical work on the application of machine learning-based technologies to fight COVID-19 and new perspective to solve the identified problems are presented in this paper. Furthermore, we believed that the presented survey with bibliometric analysis could make it easier for researchers to identify areas that need further development and possibly identify potential collaborators at author, country and institutional level, with the overall aim of furthering research in the focused area of machine learning application to disease control.

1   # Early survey with bibliometric analysis on machine learning approaches in
2   # controlling coronavirus

4   Haruna Chiroma[1], Absalom E. Ezugwu[2,*], Fatsuma Jauro[3], Mohammed A. Al-Garadi[4], Idris N. Abdullahi[5], and
5   Liyana Shuib[6]

6   [1]Future Technology Research Center, National Yunlin University of Science and Technology, Douliu, Taiwan
7   [2]School of Computer Science, University of KwaZulu-Natal, Pietermaritzburg, KwaZulu-Natal South Africa
8   [3]Department of Computer Science, Faculty of Science, Ahmadu Bello University, Zaria, Nigeria
9   [4]Department of Biomedical Informatics, Emory University, Atlanta, USA
10   [5]Department of Medical Laboratory Science, College of Medical Sciences, Ahmadu Bello University, Zaria, Nigeria
11   [6]Department of Information System, University of Malaya, Kuala Lumpur, Malaysia

13   **Corresponding Authors**: Absalom E. Ezugwu, School of Computer Science, University of KwaZulu-Natal,
14   Pietermaritzburg, KwaZulu-Natal, South Africa. Email: ezugwua@ukzn.ac.za

16   **Abstract**

17   **Background and Objective:** The COVID-19 pandemic has caused severe mortality across the globe with the USA
18   as the current epicenter, although the initial outbreak was in Wuhan, China. Many studies successfully applied machine
19   learning to fight the COVID-19 pandemic from a different perspective. To the best of the authors' knowledge, no
20   comprehensive survey with bibliometric analysis has been conducted on the adoption of machine learning for fighting
21   COVID-19. Therefore, the main goal of this study is to bridge this gap by carrying out an in-depth survey with
22   bibliometric analysis on the adoption of machine-learning-based technologies to fight the COVID-19 pandemic from
23   a different perspective, including an extensive systematic literature review and a bibliometric analysis.

24   **Methods:** A literature survey methodology is applied to retrieve data from academic databases, and a bibliometric
25   technique is subsequently employed to analyze the accessed records. Moreover, the concise summary, sources of
26   COVID-19 datasets, taxonomy, synthesis, and analysis are presented. The convolutional neural network (CNN) is
27   found mainly utilized in developing COVID-19 diagnosis and prognosis tools, mostly from chest X-ray and chest
28   computed tomography (CT) scan images. Similarly, a bibliometric analysis of machine-learning-based COVID-19-
29   related publications in Scopus and Web of Science citation indexes is performed. Finally, a new perspective is
30   proposed to solve the challenges identified as directions for future research. We believe that the survey with
31   bibliometric analysis can help researchers easily detect areas that require further development and identify potential
32   collaborators.

33   **Results:** The findings in this study reveal that machine-learning-based COVID-19 diagnostic tools received the most
34   considerable attention from researchers. Specifically, the analyses of the results show that energy and resources are
35   more dispensed toward COVID-19 automated diagnostic tools, while COVID-19 drugs and vaccine development
36   remain grossly underexploited. Moreover, the machine-learning-based algorithm predominantly utilized by
37   researchers in developing the diagnostic tool is CNN mainly from X-rays and CT scan images.

38   **Conclusions:** The challenges hindering practical work on the application of machine-learning-based technologies to
39   fight COVID-19 and a new perspective to solve the identified problems are presented in this study. We believe that
40   the presented survey with bibliometric analysis can help researchers determine areas that need further development
41   and identify potential collaborators at author, country, and institutional levels to advance research in the focused area
42   of machine learning application for disease control.

43   **Keywords**: machine learning, bibliometric analysis, CT scan, convolutional neural network, COVID-19 diagnosis
44   tool

1. Introduction

The novel 2019 coronavirus disease pandemic emerged on December 12, 2019, and the Chinese government announced the isolation of new types of coronavirus on January 7, 2020 (Imai et al., 2020). This virus was code-named as COVID-19 by the World Health Organization (WHO) on January 12, 2020. However, it was renamed to Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) by the international committee on the taxonomy of the virus (Gorbalenya, 2020). Since the pandemic, the number of confirmed cases as of October 10, 2020 reached 36,616,555 with 1,063,429 recorded fatalities. According to the origin of the first confirmed case report, the infection's transmission was from animal to human, that is, zoonotic agent (Read et al., 2020). The virus has extended beyond China to other continents of the world (Wang et al., 2020). The rise in the number of cases in Wuhan and other countries after the evacuation of the cases and closure of the marketplace in Wuhan shows a secondary transmission between humans. Related to the severe acute respiratory syndrome (SARS), the COVID-19 pandemic occurred during the China spring festival, the most famous celebration in China that attracted about 3 million Chinese people who traveled throughout the country. This festival period might have created an avenue for the transmission of this contiguous disease, making its prevention and control very difficult. The report on the secondary cases was made approximately 10 days after the first outbreak of the virus. These new patients were infected through human-to-human transmission because they never had any contact with the Wuhan marketplace, the origin of the first case of the pandemic (Read et al., 2020). The first non-Chinese infected case that spread over the provinces of China and then to the continents of Asia was reported on January 13, 2020, without any connection to the epidemiology at the market place in Wuhan (Hui et al., 2020). Subsequently, the spread continued, and various cases from countries abroad, such as the USA, Italy, and France began to be reported (Holshue et al., 2020).

Human-to-human transmission usually occurs in the case of close contact between humans. Restricting human-to-human spread is crucial for decreasing secondary infections among healthcare workers and close contact. The WHO recommended infection control interventions to reduce the risk of the general spread of acute respiratory diseases such as avoiding close contact with infected patients, frequent washing of hands in case of contact with ill people, and avoiding unprotected contacts with wild or farm animals. People suffering from the acute respiratory infection should always observe cough etiquette by maintaining distancing, covering their cough, sneezing with disposable tissues, and washing their hands frequently. Within the healthcare system, equipment improved standard infection control and prevention practices are recommended, especially in the hospital emergency unit (WHO, 2020). Progressively, an interim clinical guideline has been established by the US Center for Disease Control and Prevention for the COVID-19 pandemic to place control measures and reduce the spread of SARS-CoV-2 in the USA (Patel & Jernigan, 2020).

In minimizing the devastating effect of the COVID-19 pandemic and fighting the virus, the timely processing of COVID-19 data analytics is highly critical. The data can come from different aspects such as the patients, fractionated into molecular, society, and population, which can help in the prevention and treatment of COVID-19 (Park et al., 2020). Machine learning is one of the major technologies in combatting COVID-19 (Elavarasan & Pugazhendhi, 2020; Kumar et al., 2020). Much effort based on synergy from the machine learning and biomedical research community is ongoing to develop different approaches to combat the disease. Ongoing efforts include the screening of patients, detection and monitoring of patients, development of drugs, repurposing of drugs to treat patients with COVID-19, predicting the protein structure related to COVID-19, COVID-19 diagnostic systems, COVID-19 vaccine development, and analyzing personalized therapeutic effects for the evaluation of new patients (Alimadadi et al., 2020). The overall goal of these disease control techniques is to prevent the spread; track confirmed cases, recoveries, and mortality; and predict future pandemics (Vaishya et al., 2020).

Many studies adopted machine learning to fight the COVID-19 pandemic from a different perspective. A systematic literature review of published and preprinted reports on prediction models for diagnosing COVID-19 was reported. Prediction models were found floating the academic literature very rapidly to help in diagnosing COVID-19 and making critical medical decisions (Wynants et al., 2020). However, the major issue with the work is that it uses preprint publication that has not been validated by peer review. The review is not specific to machine learning for COVID-19. Sources of COVID-19 datasets that can guide researchers access different data for studies on COVID-19 are absent. The study is limited to the diagnosis and prognosis of COVID-19. A comprehensive taxonomy on the prediction tools to show the area that needs improvement is missing in the survey. Bibliometric analysis is not integrated into the

95   study. Similar bibliometric analyses have been reported in the literature as presented by Chahrour et al. (2020),
96   Hossain (2020), and Lou et al. (2020). However, these existing analyses differ from the current bibliometric analysis
97   in this study because the current analysis result focuses on the application of machine learning techniques to combat
98   COVID-19 pandemic as opposed to various literature reporting general medical practices on COVID-19.

99   In this study, we propose to conduct a dedicated comprehensive survey on the adoption of machine learning to fight
100   the COVID-19 pandemic from a different perspective, including an extensive literature review and a bibliometric
101   analysis. To the best of our knowledge, this study is the first comprehensive analysis of research output focusing on
102   several possible applications of machine learning techniques for mitigating the worldwide spread of the ongoing
103   COVID-19 pandemic. We are mindful that other publications might not be captured in our scope because the current
104   study is only limited to the eight academic databases mentioned in Table 1. We are also very dependent on the indexing
105   of the databases used, which is akin to any other bibliometric research study.

106   Other sections of the study are organized as follows: Section 2 presents the methodology for the survey. Section 3
107   presents the rudiments of the major machine learning algorithms used in fighting the COVID-19 pandemic. Section 4
108   presents the adoption of machine learning to fight COVID-19. Section 5 unravels the different sources of COVID-19
109   datasets. Section 6 discusses the survey and bibliometric analysis. Section 7 unveils challenges and future research
110   directions before the conclusion in Section 8. Figure 1 presents the visual structure of the survey paper, which is
111   similar to the work in (Mohammadi et al., 2018).

112

113
114   **Figure 1: Graphical representation of the survey structure**

115

116

117

118   **2. Methodology**

119   This section presents the protocol followed to survey the adoption of machine learning in fighting COVID-19. The
120   survey was performed based on a systematic literature review procedure in computer science given by (Weidt and
121   Silva, 2016). The search keywords, techniques, data sources, databases, and inclusion/exclusion criteria used are
122   discussed.

123   **2.1 Search keywords**

124   Initial search keywords were carefully selected based on the defined research goal. After an initial search, multiple
125   keywords were formulated from new words found in several relevant articles. The keywords were later reduced to suit
126   the objective of the research. The keywords used for the study included "deep learning, COVID-19," "convolutional
127   neural network, COVID-19," "long short term memory, COVID-19," "artificial neural network, COVID-19,"
128   "machine learning, COVID-19," "decision tree, COVID-19," "COVID-19 diagnosis tool," and "COVID-19 decision
129   support system."

130   **2.2 Data sources**

131   The defined search keywords were used to retrieve relevant journal articles and conference papers published by
132   prominent peer-reviewed journals indexed in various academic databases. Table 1 shows the different academic
133   databases used to obtain articles for the survey.

134

135

136   **Table 1: Academic databases**

137

2.3 Article inclusion/exclusion criteria

139 Inclusion/exclusion criteria were set up based on the research aim to decide which articles are eligible for the next
140 review stage. Articles that meet the inclusion criteria were considered relevant for the research, and those that do not
141 meet the inclusion criteria were excluded. The set inclusion/exclusion criteria are provided in Table 2.

142

143

144 **Table 2: Article inclusion/exclusion criteria**

145

2.4 Article selection

147 Article selection for this research followed a three-stage analysis. The first analysis stage considered only the titles
148 and abstracts of the papers to extract relevant papers. The second analysis stage considered the analysis of the abstract,
149 introduction, and conclusion to refine the selection in the first stage. At the third and final analysis stages, papers were
150 read thoroughly, and a threshold was set to rate the quality of papers in terms of their relevance to the research. A
151 paper was selected if it reported an empirical application of machine learning to fight COVID-19 similar to Rodriguez-
152 Morales et al. (2020). Articles that met the threshold value were selected, and those below the threshold were dropped.
153 Figure 2 shows the total number of papers obtained from the academic databases and the final number of papers
154 considered for the research after applying all the extraction criteria.

155
156
157
158 **Figure 2: Article selection**

159

160

2.5 Bibliometric protocol

162 VOSviewer software was used to present a bibliometric analysis of the existing literature on COVID-19. VOSviewer
163 software is a tool for constructing and visualizing bibliometric maps of items, such as journals, research, or individual
164 publications. These maps can be created based on citation, bibliographic coupling, co-citation, or co-authorship
165 relations. The bibliometric analysis software also offers text mining functionality that can be used to construct and
166 visualize co-occurrence networks of important terms extracted from a body of scientific literature (see
167 www.vosviewer.com). We only used 1,178 publications with the keyword "novel coronavirus" and 98 publications
168 with the keyword "COVID-19 and artificial intelligence" that were retrieved from Scopus and Web of Science
169 academic databases for the bibliometric analysis presented in this study. Only 57 document results were extracted
170 using the keyword "COVID-19 and machine learning" from the same academic database. Articles from other sources
171 were not considered because most of these publications have not been peer-reviewed and are not available online in
172 the form of preprint publications.

3. Theoretical background of machine learning algorithms

174 Numerous machine learning algorithms exist in the literature. In this section, we discuss the basic operations of the
175 major machine learning algorithms used in fighting the COVID-19 pandemic to provide the readers with a basic
176 concept of how the algorithms operate in achieving their goal, especially novice researchers, thus making the study
177 self-contained. The main algorithms include artificial neural network (ANN), convolutional neural network (CNN),
178 and long short term memory (LSTM).

3.1 Artificial neural network

ANN is a machine learning algorithm designed to emulate the human brain. Similar to neurons in the human brain, ANN consists of interconnected nodes. A given ANN consists of three essential components, namely, node (neuron) character, network topology, and several learning rules (Livingstone, 2008). Signal processing strategy involving the number of inputs and outputs associated with a node, the weight of each input and output, and the activation function are all determined by the node character. The organization of the nodes and the interconnection between them are determined by the network architecture, which usually consists of three underlying layers (input, output, and hidden layers). Learning is employed to train the network. The learning rules decide on weight initialization and adjustment. An ANN is said to have learnt if it can process probabilistic, fuzzy, noisy, and vague data with an insignificant negative effect on the output quality, and generalize acquired knowledge to unknown tasks (Basheer & Hajmeer, 2000). A basic ANN model consists of multiple nodes such that every node receives multiple inputs from other nodes through connections having associated weights. The weighted sum of the inputs is passed through a threshold gate. The node is activated and transmits the output to another node only if the weighted sum of its input exceeds the threshold (Basheer & Hajmeer, 2000).

3.2 Convolutional neural network

CNN is the most commonly used deep learning algorithm (Haque & Neubert, 2020). It is a discriminative deep learning algorithm formed by a stack of multiple convolutional and pooling layers (Deng, 2014). The major strength of CNN lies in its ability to perform parameter sharing, sparse interaction, and equal representation. The network is faster and easier to train due to its utilization of local connections and sharing of weights (Pouyanfar et al., 2018). A typical CNN receives an image as its input and has neurons arranged in a 3D form connecting to only a portion of the previous layer (Haque & Neubert, 2020). The architecture of CNN is presented in series of stages, where the first stages consist of convolution and pooling layers, and the final stage is composed of a fully connected layer (Lecun, Bengio, & Hinton, 2015).

CNN receives 3D input x in the form $m \times m \times r$, where m is the height and width of the input, and r is the channel number, for example in RGB $r = 3$. The convolution layer has several filters known as kernels k with a size of $n \times n \times q$ where $n < x$ and $q \leq r$. The kernels share the same parameters of weight (W^k) and bias (b^k). During convolution, k feature maps are generated $(h^k h)$, each having size $m - n - 1$. The convolution layer finds the dot product between w and input x, and applies activation function f to the output expressed as (Pouyanfar et al., 2018): $h^k = f(W^k * x + b^k)$. The pooling layer prevents overfitting and hastens training by reducing the length of the feature map (Pouyanfar et al., 2018; R. Zhao et al., 2019). Maximum and average pooling are the widely applied pooling methods. Finally, the fully connected SoftMax layers are deployed for prediction (Zhao et al., 2019).

3.3 Long short term memory

The LSTM deep learning network is a type of recurrent neural network that can recall data features over several intervals of time (Liu et al., 2019). It was developed to solve the vanishing gradient weakness of the recurrent neural network (Karim, Majumdar, Darabi, & Chen, 2018). The hidden layers of LSTM are treated as memory cells, making the network powerful in handling long- and short-term correlation in a time series (Zhao et al., 2017). The earliest version of LSTM contained memory cells, and input and output gates with no forget gate; the forget gate was later deployed to enable continual learning of tasks by resetting the state of the LSTM (Greff et al., 2016). Its updated architecture consists of multiple LSTM units with each unit having an input gate, a forget gate, an output gate, and a memory cell. Sak et al. (2014) described the underlying architecture of LSTM as consisting of memory blocks in its hidden layer. The memory blocks have memory cells for the storage of the temporal state of the network with additional units, known as gates, to supervise the information flow. A memory cell has an input gate that manages the inflow of input activations to the memory cell and an output gate to manage the outflow of cell activations to the network. The forget gate is incorporated to forget or reset the memory of the cell adaptively. LSTM computes mapping iteratively at a timestamp $t = 1\ to\ T$ with an input sequence $x = (x_1, x_2,...,x_r)$ and an output sequence $y = (y_1, y_2,...,y_r)$. LSTM is good at addressing complex sequential machine learning problems (Karim et al., 2018). Deep LSTM

225 architectures consist of stacked LSTM layers (Sak et al., 2014). LSTMs are strong in handling temporal dependencies
226 in sequences but weak in dealing with long sequence dependencies (Karim et al., 2018).

4. COVID-19 machine learning adoption-oriented perspective

228 In the fight against COVID-19, different aspects of artificial intelligence (AI) were applied to curtail its adverse effect
229 (Dananjayan & Raj, 2020). The taxonomy in Figure 3 was created from the project that involved machine learning in
230 fighting COVID-19. The data used in creating the taxonomy were extracted from the papers that applied the machine
231 learning algorithm to fight COVID-19.

232

233 **Figure 3: Taxonomy of the machine learning algorithms adopted in fighting COVID-19**

234

4.1 COVID-19 diagnostic tools

236 Currently, the sensitivities for reverse transcription-polymerase chain reaction (RT–PCR)-based viral nucleic acid
237 assay are used as the reference standard method to confirm COVID-19 infection (Corman et al., 2020). However, such
238 a laboratory test is time consuming, and the supply of test kits may be the bottleneck for a rapidly growing suspicious
239 population even for many developed countries such as the USA. More importantly, initial false-negative or weakly
240 positive RT–PCR test results were found in several later-confirmed cases, while highly suspicious computed
241 tomography (CT) imaging features were present (Xu et al., 2020; Xie et al., 2020). The treatment and screening of
242 COVID-19 can be more effective when deep learning approach, CT features, and real-time RT–PCR results are
243 integrated (Li et al., 2020). AI and deep learning can assist in developing diagnostic tools and deciding on treatment
244 (Rao and Vazquez, 2020; Shi et al., 2020). As a result, many diagnostic tools were developed based on the machine
245 learning algorithm to fight COVID-19. For example, Apostolopoulos and Mpesiana (2020) applied transfer learning
246 with CNN to detect COVID-19 from X-ray images containing common bacterial pneumonia and normal incidents and
247 established COVID-19 infection. Transfer learning CNN was used to diagnose COVID-19 cases from X-ray datasets.
248 The results indicated that VGG19 diagnosed COVID-19 confirmed cases with better accuracy on two- and three-
249 classification problems compared with MobileNet v2, Inception, Xception, and Inception ResNet v2. The proposed
250 approach can help develop a cost-effective, fast, and automatic COVID-19 diagnostic tool, and reduce the exposure
251 of medical workers to COVID-19. Similarly, Rahaman et al. (2020) developed an automated computer-aided diagnosis
252 (CAD) system for the detection of COVID-19 samples from healthy cases and cases with pneumonia using chest X-
253 ray (CXR) images. Their study demonstrated the effectiveness of applying deep transfer learning techniques for the
254 identification of COVID-19 cases using CXR images.

255 Ardakani et al. (2020) were motivated by the time consumption and high cost of the traditional medical laboratory
256 COVID-19 test to investigate the performance of 10 well-known CNNs in diagnosing COVID-19. The 10 variants of
257 CNN included VGG-16, AlexNet, Xception, VGG-19, ResNet-101, SqueezeNet, ResNet-50, GoogleNet, MobileNet-
258 V2, and ResNet-18. All the CNN variants were applied on CT scan images because the CT slice is a fast method of
259 diagnosing patients with COVID-19. The diagnostic results of the CNN variants indicated that ResNet-101 and
260 Xception outperformed the other CNN variants in diagnosing COVID-19. They concluded that ResNet-101 has a high
261 sensitivity in characterizing and diagnosing COVID-19 infections. Therefore, it can be used as an alternative tool in
262 the department of radiology for diagnosing COVID-19 infection. It is cheaper and faster compared with traditional
263 laboratory analysis.

264 Butt et al. (2020) applied CNN for the detection of COVID-19 from the chest CT scan of patients. CNN was found
265 very fast and reliable in the detection of COVID-19 from a chest CT scan compared with the conventional RT–PCR
266 testing. In summary, the CNN model is fast and reliable in detecting COVID-19 infection. Huang et al. (2020) applied
267 a deep learning algorithm on a chest CT scan of a patient with COVID-19 to quantify lung burden changes. The
268 patients with COVID-19 were grouped into mild, moderate, severe, and critical based on findings from the chest CT
269 scan, clinical evaluation, and laboratory results. Deep learning algorithm was applied to assess the lung burden
270 changes. They found that the assessment of lung opacification measured on the chest CT scan substantially differed

271 from that of the clinical groups. The approach can remove the subjectivity in the initial assessment of COVID-19
272 findings.

273

274 **Table 3: Summary of COVID-19 diagnostic tools based on machine learning algorithms**

275

276 Mei et al. (2020) proposed a joint model comprising CNN, support vector machine (SVM), random forest (RF), and
277 multilayer perceptron integrated with chest CT scan result and non-image clinical information to predict COVID-19
278 infection in a patient. CNN was run on the CT image, while the other algorithms classified COVID-19 using the non-
279 image clinical information. The output of the CNN and the different algorithms were combined to predict the patient's
280 COVID-19 infection. The diagnostic tool can rapidly detect COVID-19 infection in patients. Liu et al. (2020) used
281 logistic regression for the prediction of COVID-19 infection sliding to the severity of the COVID-19 cohort. The
282 results of the study showed that the CT quantification for the pneumonia lesions could predict the progression of a
283 patient with COVID-19 to a severe stage at an early, non-invasive level. This situation can provide a prognostic
284 indicator for coronavirus clinical management. Jiang et al. (2020) applied a machine learning algorithm to predict
285 COVID-19 clinical severity. They developed a predictive tool that predicts patients at risk for increased COVID-19
286 severity at the first presentation. The survey can help in the optimal utilization of scarce resources to cope with the
287 COVID-19 pandemic.

288 Hurt et al. (2020) collected CXR images from patients with COVID-19 in China and America. They applied a deep
289 learning algorithm for the early diagnosis of COVID-19 from the CXR. They found that deep learning predicted and
290 consistently localized areas of pneumonia. The deep learning algorithm can diagnose a patient's COVID-19 infection
291 early. Loey et al. (2020) were motivated by the insufficient COVID-19 dataset to propose a generative adversarial
292 network (GAN) and CNN variant to detect COVID-19 in patients. GAN was used to generate more X-ray images.
293 Googlenet, Alexnet, and ResNet18 were applied as the deep transfer learning models. They found that Googlenet and
294 Alexnet scored 80.6%, 85.2%, and 100%, respectively, in the four-, three-, and two-class classification problem,
295 respectively. The study's method can facilitate the early detection of COVID-19 and reduce the workload of a
296 radiologist. Wu et al. (2020) proposed a multi-view ResNet50 for the screening of COVID-19 from chest CT scan
297 images. ResNet50 was trained with the multi-view chest CT scan images. The results showed that the multi-view
298 ResNet50 fusion achieved a high performance compared with the single view. The diagnosis tool developed can
299 reduce the workload of a radiologist by offering fast, accurate COVID-19 diagnosis. Ucar and Korkmaz (2020)
300 developed a rapid COVID-19 diagnosis tool from X-ray images based on SqeezeNet (a pre-defined CNN) and the
301 Bayesian optimization method. The SqueezeNet hyperparameters were optimized using the Bayesian optimization
302 method. Bayesian optimization-based SqueezeNet was applied to detect COVID-19 from X-ray images labeled
303 normal, pneumonia, and COVID-19. Bayesian-based SqueezeNet outperformed the baseline diagnostic tools.

304 Togaçar et al. (2020) applied CNN for the exploitation of social mimic and CXR based on fuzzy color and the stacking
305 method to diagnose COVID-19. The stacked data were trained using CNN, and the features obtained were processed
306 with mimicking social optimization. The compelling features were used for classification into COVID-19, pneumonia,
307 and standard X-ray imagery using SVM. Singh et al. (2020) used CNN and multi-objective differential evolution
308 (MODE) for the early detection of COVID-19 from a chest CT scan image. The initial parameters of the CNN were
309 tuned using MODE to create a MODE-based CNN and classify patients with COVID-19 based on positive or negative
310 chest CT scan images. MODE-based CNN outperformed the competitive models (ANN, ANFIS, and traditional
311 CNN). The proposed method is beneficial for COVID-19 real-time classification owing to its speed in diagnosing
312 COVID-19. Salman et al. (2020) constructed a CNN-based COVID-19 diagnostic tool for the detection of COVID-
313 19 from CXR images. CNN–inceptionV3 was applied to detect COVID-19 from 130 X-ray images of patients infected
314 with COVID-19 and 130 normal X-ray images. The results indicated that CNN–inceptionV3 could detect COVID-19
315 from the X-ray images and reduce the testing time required by a radiologist.

316 Ozturk et al. (2020) used CNN to develop an automated tool for diagnosing COVID-19 from raw CXR images. Binary
317 and multi-class categories were experimented on using a CNN with 17 convolution layers with a different filter on

318   each convolution layer. The model can be used for the early screening of patients with COVID-19 and assist the
319   radiologist in validating COVID-19 screening. Li et al. (2020b) developed an automated framework based on CNN
320   for the detection of COVID-19 from chest CT scan and differentiate it from community-acquired pneumonia. The
321   study collected data from 3,322 patients comprising 4,356 chest CT scans. CNN was applied to detect patients with
322   COVID-19 and typical community pneumonia. The experiment results showed that CNN can distinguish patients with
323   COVID-19 from those with community-acquired pneumonia and other similar lung diseases. The proposed framework
324   automated the COVID-19 testing and reduced the testing time and fatigue. Yang et al. (2020b) applied densely
325   connected convolutional networks optimized with stochastic gradient descent algorithm for the detection of COVID-
326   19 from chest CT scan images. Oh et al. (2020) applied patch-based CNN–ResNet-18 (P-CNN) due to lack of
327   sufficient training data for diagnosing COVID-19 from CXR images. The study used imaging biomarkers of the CXR
328   radiographs. P-CNN ResNet-18 was applied, and P-CNN produced clinically salient maps that are useful in diagnosing
329   COVID-19 and patient triage. P-CNN ResNet-18 achieved the best result compared with the baselined algorithm
330   performance. The limited amount of data can be used for COVID-19 diagnoses and were interpretable. Table 3
331   summarizes the diagnostic tools developed based on machine learning. Refer to Dong et al. (20202) for an engaging
332   research review on the role of imaging in the detection and management of COVID-19 disease spread.

4.2 COVID-19 decision support system

334   The decision support system related to COVID-19 can help decision/policymakers formulate policy to curtail COVID-
335   19. Many COVID-19 decision support systems were developed based on machine learning approaches. For example,
336   Ayyoubzadeh et al. (2020) applied LSTM and linear regression to predict the number of positive cases in Iran. LSTM
337   and linear regression were used on Google search data to predict the COVID-19 cases in Iran. The results indicated
338   that linear regression outperforms LSTM in predicting the positive cases of COVID-19. The algorithm can predict the
339   trend of the COVID-19 pandemic in Iran, which can help policymakers plan the allocation of medical resources.
340   Chimmula and Zhang (2020) applied deep LSTM for forecasting COVID-19 transmission and possible COVID-19
341   ending period in Canada and other parts of the world. The transmission rate of Canada was compared with that of Italy
342   and the USA. The future outbreak of the COVID-19 pandemic was predicted to help Canadian decision makers
343   monitor the COVID-19 situation and prevent the future transmission of the epidemic in Canada.

344   Liu et al. (2020b) proposed ANN in modeling the trend of COVID-19 and restoring the operational capability of
345   medical services in China. ANN was used for modeling the pattern of COVID-19 in Wuhan, Beijing, Shanghai, and
346   Guangzhou. Autoregressive Integrated Moving Average (ARIMA) was applied for the estimation of nonlocal hospital
347   demands for the period of COVID-19 pandemic in Beijing, Shanghai, and Guangzhou. The results indicated that the
348   number of people infected with COVID-19 would increase by 45%, while death would increase by 567%. COVID-
349   19 will reach its peak by March 2020 and toward the end of April 2020. This finding will assist policymakers and
350   health officials in planning to deal with challenges of the unmet medical requirement of other diseases during the
351   COVID-19 pandemic.

352

353   **Table 4: Summary of the adoption of machine learning approaches in building COVID-19 decision support**
354   **systems**

355

356   Pirouz et al. (2020) proposed a group method of data handling in a neural network to predict the number of COVID-
357   19 confirmed cases based on weather conditions. The dominant weather condition used included temperature, city
358   density, humidity, and wind speed. The results indicated that humidity and temperature have a substantial influence
359   on COVID-19 confirmed cases. Temperature and humidity influence COVID-19 negatively and positively,
360   respectively. These results can be used by decision makers to manage the COVID-19 pandemic. Yang et al. (2020b)
361   applied LSTM to predict the COVID-19 trend in China. The prediction model indicated that the COVID-19 pandemic
362   should peak toward the end of February 2020 and start declining at the end of April 2020. The prediction model can
363   be used by authorities in China to decide in controlling the COVID-19 pandemic. Vaid et al. (2020) adopted a machine
364   learning approach to predict COVID-19 potential infections based on reported cases in North America. Critical

365 parameters were identified from dimension reduction. Passed diseases were inferred from recent fatalities using a
366 hierarchical Bayesian estimator. The model predicted potential COVID-19 infections in North America. Policymakers
367 in North America can use the projection to curtail the effect of the COVID-19 pandemic. Tuli et al. (2020) developed
368 a machine learning COVID-19 predictive model and deployed it in the cloud computing environment for real-time
369 tracking of COVID-19, predicting the growth and potential thread of COVID-19 in different countries worldwide.
370 Government and citizens can use the results for proactive measures to fight COVID-19.

371 Tiwari et al. (2020) used a machine learning approach to predict the COVID-19 pandemic number of cases, recoveries,
372 and deaths in India based on data from China. The prediction results indicated that COVID-19 would peak between
373 the third and fourth week of April 2020. The Indian government can use the study to formulate policies and decide on
374 mitigating the spread of COVID-19. Ribeiro et al. (2020) evaluated six machine learning algorithms, namely, Cubis
375 regression (CUBIST), RF, ridge regression (RIDGE), support vector regression (SVR), ARIMA, and stacking-
376 ensemble learning (SEL), on COVID-19 datasets collected in Brazil to predict confirmed cases for one, three, and six
377 days ahead. They found that SVR outperformed RIDGE, ARIMA, RF, CUBIST, and SEL. The study can help monitor
378 COVID-19 cases in Brazil and facilitate critical decisions on COVID-19. Tummers et al. (2020) applied k-means to
379 cluster documents based on COVID-19 and people with intellectual disability. Table 4 summarizes the studies on
380 COVID-19 decision support system.

4.3 COVID-19 from genome sequences

382 The protein sequence of COVID-19 can be collected to apply the machine learning approach for the prediction of
383 COVID-19 (Qiang et al., 2020). For example, Qiang et al. (2020) predicted the infection risk of non-human origin of
384 COVID-19 from spike protein for prompt alarm using RF. The genome data comprised of non-human COVID-19
385 origin (positive) and human COVID-19 origin (negative). RF was applied for the training to predict non-human
386 COVID-19 origin. The results showed that the RF model achieved high accuracy in predicting non-human COVID-
387 19 origin. The study can be used in COVID-19 genome mutation surveillance and exploring evolutionary dynamics
388 in a simple, fast, and large-scale manner. Randhawa et al. (2020) combined decision tree and digital signal processing
389 (DT–DSP) to detect the COVID-19 virus genome and identified the signature of intrinsic COVID-19 viruses' genome.
390 DT–DSP was applied to explore over 5,000 viral genome sequences with 61.8 million by the 29 viral sequences of
391 COVID-19. The result obtained supported the bat origin of COVID-19 and successfully classified COVID-19 with
392 100% accuracy as sub-genus *Sarbecovirus* within *Betacoronavirus*. DT–DSP is a reliable real-time alternative
393 taxonomic classification. Table 5 summarizes the studies.

394

395 **Table 5: Summary of detecting COVID-19 from genome sequence via an algorithm**

396

4.4 COVID-19 drug discovery

398 Machine learning and AI provide approaches for the speedy processing of a large amount of collected medical data
399 generated daily as well as the extraction of new information from transversely different applications. In the prediction
400 of disease, a viral mutation can be forecast before the emergence of new strains. It also allows the prediction of new
401 structure and availability of broader structural information. Efficient drug repurposing can be achieved in mining
402 existing data. The stages for the development of COVID-19 drugs are as follows (Park et al., 2020): *Disease
403 prediction*: The prediction of future-generation viral mutation can be accomplished by AI and machine learning
404 approaches. *Structural analysis*: The COVID-19 structure and primary functional site are characterized. *Drug
405 repurposing*: For insight into new disease treatment, existing drug data are mined. *New drug development*: Efficiencies
406 across the entire pharmaceutical life cycle are achieved by rapid processing. Ke et al. (2020) applied machine learning
407 to identify drugs already marketed that can treat COVID-19. They compiled two independent datasets to develop two
408 machine learning models. The first model was built based on drugs that are known to have antiviral activities. The
409 second model was built based on 3C-like protease inhibitors. The database of market-approved drugs was screened
410 by the machine learning model to predict the drugs with potential antiviral activities. The drugs predicted to have
411 antiviral activities were evaluated against the antiviral activities by a cell-based feline infectious peritonitis virus

412 duplication assay. The assay results were the machine learning model feedbacks for incremental learning of the model.
413 Finally, 80 marketed drugs were identified to have potential antiviral activities. Old drugs with antiviral activities
414 against feline infectious peritonitis COVID-19 were found.

415 **4.5 COVID-19 vaccine development**

416 Typically, the immune system is prepared to elicit antibody or cell-mediated responses against a pathogen by a vaccine
417 that protects the body from infectious diseases. Immunogenicity is the vaccine ability to the response. For a long-time,
418 effective immunity, the vaccine has to properly activate innate, adaptive responses (Klein et al., 2010).

419 The following phases should be adopted to develop a COVID-19 vaccine (Gonzalez-Dias et al., 2020): *Dataset*
420 *preparation*: The quality of the data to be used influences the machine learning algorithm. Thus, preparing quality
421 data before feeding into the algorithm is sacrosanct. Data come in different sizes ranging from small, medium, and
422 large. Data quality must be ensured because a quality immune response is needed. The reliability of the data needs to
423 be guaranteed by ensuring that the serological assay is well qualified in case it is not validated based on known
424 parameters (linearity, specificity, LLOQ, ruggedness, LLOD, ULOQ, and reproducibility). *Vaccines and relevant*
425 *genes*: In vaccinology, the machine learning algorithm is trained to discover the combination of genes and the best
426 vaccines parameters. The data for the training are extracted from omics experiment, which will be used to obtain the
427 required combination. Feature selection is performed to find the best representative of the discriminatory gene
428 signatures. Then, the new vaccines are predicted. The three main feature selection methods are filter, wrapper, and
429 embedded. *Machine learning algorithm selection*: This task is not a straightforward task because many factors must
430 be considered before selecting the appropriate algorithm for the modeling. The choice of the algorithm depends on
431 the nature of the data, and the options include supervised, unsupervised, and semi-supervised learning. For instance,
432 if the data have no output, then unsupervised learning algorithm, e.g., k–the nearest neighbor is the possible candidate
433 algorithm for the modeling but is not guaranteed. Many algorithms have to be tested on solving the same problem
434 before the algorithm that produces the best output is selected. *Model testing*: The performance of the model is tested.
435 The data are partitioned into training and testing; the former is used for training the algorithm, and the latter is used
436 for evaluating the performance of the model using several performance parameters, e.g., MSE, accuracy, and F-
437 measure (Gonzalez-Dias et al., 2020). The application of a machine learning algorithm to sift through trillions of
438 compounds of the vaccine adjuvants can shorten the vaccine development time. The machine learning algorithm can
439 be used for screening compounds for a potential adjuvant candidate for the SARS-CoV-2 vaccine (Ahuja et al., 2020).

440 **5. COVID-19 datasets**

441 **5.1 Data sources**

442 Ahuja et al. (2020) reported that COVID-19 data are now growing. In this section, we present the sources of COVID-
443 19 data to the machine learning community. Given the novelty of the virus, centralizing the collection of sources of
444 data will help researchers access different types of COVID-19-related data and provide them opportunities to work on
445 a different aspect of COVID-19 that may lead to novel discoveries. Table 6 has five columns, where the first, second,
446 third, fourth, and fifth columns represent the reference, data, owners, source/accessibility, and remarks, respectively.
447 We only present the projects that revealed and fully discussed their data sources.

448

449

450 **Table 6 Summary of COVID-19 data sources and accessibility**

451

452 **5.2 COVID-19 on CXR and CT scan images**

453 In this section, we discuss the diagnosis of COVID-19 based on X-ray and CT scan images because of their high value
454 in COVID-19 screening. Table 6 shows that researchers heavily utilize X-rays and CT scans in developing a machine-
455 learning-based COVID-19 diagnosis tool. Guan et al. (2020) and Wong et al. (2020) found that portable chest

456   radiography (CXR) has a sensitivity of 59% for the initial detection of COVID-19-related abnormalities. Radiographic
457   abnormalities, when present, mirror those of CT, with a bilateral lower zone, a peripherally predominant consolidation,
458   and hazy opacities (Wong et al., 2019). The radiological findings of COVID-19 on CXR are those of atypical
459   pneumonia or organizing pneumonia (Kooraki et al., 2020). Although chest CT scans are reportedly less sensitive than
460   CXRs, chest radiography remains the first-line imaging modality of choice used for patients with suspected COVID-
461   19 infection (Hui et al., 2020) because it is cheap and readily available, and can easily be cleaned. For ease of
462   decontamination, the use of portable radiography units is preferred. Chest radiographs are often normal in early or
463   mild disease. According to a recent study of patients with COVID-19 requiring hospitalization, 69% had an abnormal
464   chest radiograph at the initial time of admission, and 80% had radiographic abnormalities sometime during
465   hospitalization. The findings are reported to be most extensive about 10–12 days after symptom onset. The most
466   frequent radiographic findings are airspace opacities, whether described as consolidation or less commonly, ground-
467   glass opacity (GGO) (Wong et al., 2020). The distribution is most often bilateral, peripheral, and lower zone
468   predominant (Rodrigues et al., 2020). Unlike parenchymal abnormalities, pleural effusion is rare (3%) (Wong et al.,
469   2020). According to the Center for Disease Control (CDC), even if a chest CT or X-ray suggests COVID-19, viral
470   testing is the only specific method for diagnosis. Radiography's sensitivity was reported at only 25% for detection of
471   lung opacities related to COVID-19, among 20 patients seen in South Korea with a reported specificity of 90% (Wen
472   et al., 2020). The X-ray image should be considered a useful tool for detecting COVID-19 which is challenging the
473   healthcare system due to the overflow of patients. As the COVID-19 pandemic grinds on, clinicians on the front lines
474   may increasingly turn to radiography (Casey, 2020). The most frequent findings are airspace opacities, whether
475   described as consolidation or less commonly, GGO. The distribution is most often bilateral, peripheral, and lower
476   zone predominant (Wong et al., 2020). Much of the imaging focus is on CT. In February 2020, Chinese studies
477   revealed that chest CT achieved a higher sensitivity for the diagnosis of COVID-19 compared with initial RT–PCR
478   tests of pharyngeal swab samples (Ai et al., 2020; Fang et al., 2020). Subsequently, the National Health Commission
479   of China briefly accepted chest CT findings of viral pneumonia as a diagnostic tool for detecting COVID-19 infection
480   (Yuen et al., 2020; Zu et al., 2020).

481   The typical appearance of COVID-19 on chest CT consists of multi-lobar, bilateral, predominantly lower lung zone,
482   rounded GGOs, with or without consolidation, in a mostly peripheral distribution. However, such findings are
483   nonspecific; the differential diagnosis includes organizing pneumonia and other infections, drug reactions, and other
484   inflammatory processes. Consequently, using CT to screen for COVID-19 may result in false positives. Moreover, the
485   presence of abnormalities not typically associated with COVID-19 infection, including pure consolidation, cavitation,
486   thoracic lymphadenopathy, and nodules suggests a different etiology (Bernheim et al., 2020). COVID-19-related chest
487   CT abnormalities are more likely to appear after symptom onset, but they may also precede clinical symptoms. In a
488   retrospective study by Bernheim et al. (2020), 44% of patients presenting within two days of symptom onset had an
489   abnormal chest CT, while 91% presenting within 3–5 days and 96% presenting after six days had abnormal chest CTs.
490   Shi et al. (2020) found GGOs in 14 of 15 asymptomatic healthcare workers with confirmed COVID-19. Similarly,
491   54% of 82 asymptomatic passengers with COVID-19 on the Diamond Princess cruise ship had findings of viral
492   pneumonia on the CT (Inui et al., 2020).

493   In a prospective study by Wang et al., pure GGOs were the only abnormalities seen prior to symptom onset.
494   Subsequently, 28% of patients developed superimposed septal thickening 6–11 days after symptom onset (Wang et
495   al., 2020b). Architectural distortion evolving from GGOs appeared later in the disease course, likely reflecting
496   organizing pneumonia and early fibrosis. Long-term follow-up imaging is also needed to determine the sequelae of
497   SARS-CoV-2 infection. In a retrospective study by Das et al., 33% of patients who recovered from MERS-CoV
498   developed pulmonary fibrosis; a similar outcome following COVID-19 is likely (Das et al., 2020). Lung ultrasound
499   offers a low-cost, point-of-care evaluation of the lung parenchyma without ionizing radiation. The modality is
500   especially useful in resource-limited settings (Stewart et al., 2020). Peng et al. (2020) found that sonographic findings
501   in patients with COVID-19 correlated with typical CT abnormalities. The predominantly peripheral distribution of
502   lung involvement facilitated sonographic visibility. Characteristic findings included thickened, irregular pleural lines,
503   B lines (edema), and eventual appearance of A lines (air) during recovery. Peng et al. (2020) suggested that ultrasound
504   may be useful in recruitment maneuver monitoring and guide prone positioning.

505   Chest CT is an indispensable tool for the early screening and diagnosis of suspected COVID-19 infection in patients.
506   Previous studies confirmed that the majority of patients infected with COVID-19 exhibited common chest CT
507   characteristics, including GGOs and consolidation, which reflect lesions affecting multiple lobes or infections in the

508   bilateral lung parenchyma. Increasing evidence suggests that these chest CT characteristics can be used to screen
509   suspected patients and serve as a diagnostic tool for COVID-19-caused acute respiratory diseases (ARDS) (Xu et al.,
510   2020). These findings have led to the modification of the diagnosis and treatment protocols of SARS-CoV-2-caused
511   pneumonia to include patients with characteristic pneumonia features on chest CT but negative RT–PCR results in
512   severe epidemic areas such as Wuhan City and Hubei Province (Liu et al., 2020). Patients with negative RT–PCR but
513   positive CT findings should be isolated or quarantined to prevent clustered or wide-spread infections.
514   The critical role of CT in the early detection and diagnosis of COVID-19 becomes more publicly acceptable. However,
515   several studies also reported that a proportion of RT–PCR-positive patients, including several severe cases, had
516   initially normal CXR or CT findings (Fang et al., 2020). According to the diagnostic criteria of COVID-19, patients
517   might have no or atypical radiological manifestations even at the mild or moderate stages because several lesions are
518   easily missed in the low-density resolution of CXR, suggesting that chest CT may be a better modality with a lower
519   false-negative rate. Another possible explanation is that in several patients, the targeted organ of COVID-19 may not
520   be the lung. Multiple-organ dysfunctions, including ARDS, acute cardiac injury, hepatic injury, and kidney injury,
521   have been reported during COVID-19 infection (Huang et al., 2020b). Studies also reported the chest CT appearances
522   in patients with COVID-19 after treatment, suggesting its critical role in treatment evaluation and follow up. For
523   example, a study investigated the change in chest CT findings associated with COVID-19 at different time points
524   during the infection course (Pan et al., 2020). The results showed that most apparent abnormalities on the chest CT
525   were still observable for 10 days but disappeared at 14 days after the initial onset of symptoms. Unexpectedly, a case
526   report showed pre- and post-treatment chest CT findings of a 46-year-old woman whose RT–PCR result became
527   negative, while pulmonary lesions were reversal (Duan et al., 2020).

528

5.3. Deep learning applications for COVID-19

530   Singh et al. (2020) developed a deep CNN, which was applied in the automated diagnosis and analysis of COVID-19
531   in infected patients to save the time and energy of medical professionals. They tuned and used the hyperparameters of
532   CNN by using multi-objective adaptive differential evolution (MADE). Further in the course of their experiments
533   which were extensively carried out, they used several benchmark COVID-19 datasets. The data used to evaluate the
534   performance of their proposed model were divided into training and testing datasets. The training sets were used to
535   build the COVID-19 classification model. Then, the hyperparameters of the CNN model were optimized on the
536   training sets by using the MADE-based optimization approach. The results from the comparative analysis showed that
537   their proposed method outperformed existing machine learning models such as CNN, GA-based CNN, and PSO-based
538   CNN in terms of different metrics (including F-measure, sensitivity, specificity, and Kappa statistics).

539   Jaiswal et al. (2020) applied deep learning models for the diagnosis and detection of COVID-19, and it was called
540   DenseNet201-based deep transfer learning (DTL). The authors used these pre-trained deep learning architecture as
541   automation tools to detect and diagnose COVID-19 in chest CT scans. The DTL model was used to classify patients
542   as COVID-19 positive (+ve) or COVID-19 negative (−ve). The proposed model was also utilized to extract several
543   features by adopting its own learned weights on the ImageNet dataset along with a convolutional neural structure.
544   Extensive analysis of the experiments showed that the proposed DTL-based COVID-19 model was superior to
545   competing methods. The proposed DenseNet201 model achieved a 97% accuracy compared with other models and
546   could serve as an alternative to other COVID-19 testing kits.

547   Li et al. (2020) developed a fully automated AI system to assess the severity of COVID-19 and its progression
548   quantitatively using thick-section chest CT images. The AI system was implemented to partition and quantify the
549   COVID-19-infected lung regions on thick-section chest CT images automatically. The data generated from the
550   automatically segmented lung abnormalities were compared with those of the manually segmented abnormalities of
551   two professional radiologists by using the Dice coefficient on a randomly selected subset of 30 CT scans. During
552   manual and automatic comparisons, two biomarker images were automatically computed, namely, the portion of
553   infection (POI) and the average infection HU (iHU), which were then used to assess the severity and progression of
554   the viral disease. The performance of the assessments was then compared with patients' status of diagnosis reports,
555   and key phrases were extracted from the radiology reports using the area under the receiver's operating characteristic
556   curve (AUC) and Cohen's kappa statistics. Further in their study, the POI was the only computed imaging biomarker
557   that was effective enough to show high sensitivity and specificity for differentiating the groups with severe COVID-

558   19 and groups with non-severe COVID-19. The iHU reflected the progress rate of the infection but was affected by
559   several irrelevant factors such as the construction slice thickness and the respiration status. The results of the analysis
560   revealed that the proposed deep-learning-based AI system accurately quantified the COVID-19 strains associated with
561   the lung abnormalities, and assessed the virus' severity and its corresponding progression. Their results also showed
562   that the deep learning-based tool can help cardiologists in the diagnosis and follow-up treatment for patients with
563   COVID-19 based on the CT scans.

564   Singh, Kumar, and Kaur, (2020) used a CNN to classify patients with COVID-19 as COVID-19 +ve or COVID-19
565   −ve. The initial parameters of CNN were tuned by using MODE. The authors adopted the mutation, crossover, and
566   selection operations of the differential evolution (DE) algorithm. They extracted the chest CT dataset of COVID-19-
567   infected patients and decomposed them into training and testing groups. The proposed MODE-based CNN and
568   competitive classification models were then applied to the training dataset. They compared the competitive and
569   proposed classification models by considering different fractions of the training and testing datasets. The extensive
570   analysis showed that the proposed model classified the chest CT images at reasonable accuracy rates compared with
571   other competing models, such as ANN, ANFIS, and CNS. The proposed model was also useful for COVID-19 disease
572   classification from chest CT images.

573   Asif and Yi (2020) implemented a model that automatically detected COVID-19 pneumonia in patients using digital
574   CXR images while maximizing the accuracy in detection by using deep convolutional neural networks (DCNN). Their
575   model named DCNN-based model Inception V3 with transfer learning detected COVID-19 infection in patients using
576   CXR radiographs. The proposed DCNN also provided insights on how deep transfer learning methods were used for
577   the early detection of the disease. The experimental results showed that the proposed DCNN model achieved high
578   accuracy. The proposed model also exhibited excellent performance in classifying COVID-19 pneumonia by
579   effectively training itself from a comparatively lower collection of images.

580   Hu et al. (2020) implemented a weak supervised deep learning model for detecting and classifying COVID-19
581   infection from CT images. The proposed model minimized the requirements of manual labelling of CT images and
582   accurately detected the viral disease. The model could distinguish positive COVID-19 cases from non-positive
583   COVID-19 cases by using COVID-19 samples from retrospectively extracted CT images from multiple scanners and
584   centers. The proposed method accurately pinpointed the exact position of the lesions (inflammations) caused by the
585   viral COVID-19 and potentially provided advice on the patient's severity to guide the disease triage and treatment.
586   The experimental results indicated that the proposed model achieved high accuracy, precision, and classification as
587   well as good qualitative visualization for the lesion detections.

588   Ayyoubzadeh et al. (2020) conducted a study to predict the incidence and occurrence of COVID-19 in Iran. The
589   authors obtained data from the Google Trends website (recommender systems) and used linear regression and LSTM
590   models to estimate the number of positive COVID-19 cases from the extracted data. Root mean square error and 10-
591   fold cross-validation were used as performance metrics. The predictions obtained from the Google Trend's website
592   were not very precise but could be used to build a base for accurate models for more aggregated data. Their study
593   showed that the population (Iranians) focused on the usage of hand sanitizer and handwashing practices with antiseptic
594   as preventive measures against the disease. The authors used specific keywords related to COVID-19 to extract Google
595   search frequencies and used the extracted data to predict the degree of COVID-19 epidemiology in Iran. They
596   suggested future research direction using other data sources such as social media information, people's contact with
597   the special call center for COVID-19, mass media, environmental and climate factors, and screening registries.

598   Randhawa et al. (2020) integrated supervised machine learning with digital signal processing called MLDSP for
599   genome analyses, which were then augmented by a DT approach to the machine learning component, and a
600   Spearman's rank correlation coefficient analysis for result validation. The authors identified an intrinsic COVID-19
601   virus genome signature and used it together with a machine-learning-based alignment-free approach for an ultra-fast,
602   scalable, and highly accurate classification of the COVID-19 genomes. They also demonstrated how machine learning
603   used intrinsic genomic signature to provide a rapid alignment-free taxonomic classification of novel pathogens. The
604   model accurately classified the COVID-19 virus without having *a priori* knowledge by simultaneous processing of
605   the geometric space of all relevant viral genomes. Their result analysis supported the hypothesis of a bat origin and
606   classified the COVID-19 virus as *Sarbecovirus* within *Betacoronavirus*. Also, their results were obtained through a

607 comprehensive analysis of over 5,000 unique viral sequences through an alignment-free analysis of their 2D genomic
608 signatures, combined with a DT use of supervised machine learning, and confirmed by Spearman's rank correlation
609 coefficient analyses.

610 Farhat et al. (2020) reviewed the developments of deep learning applications in medical image analysis which targeted
611 pulmonary imaging and provided insights into contributions to COVID-19. The study covered a survey of various
612 contributions from diverse fields for about three years and highlighted various deep learning tasks such as
613 classification, segmentation, and detection as well as different pulmonary pathologies such as airway diseases, lung
614 cancer, COVID-19, and other infections. The study summarized and discussed current state-of-the-art approaches in
615 the research domain, highlighting the challenges, especially given the current situation of COVID-19. First, the authors
616 provided an overview of several medical image modalities, deep learning, and surveys on deep learning in medical
617 imaging, in addition to available datasets for pulmonary medical images. Second, they provided a summarized survey
618 on deep-learning-based applications and methods on pulmonary medical images. Third, they described the COVID-
619 19 disease and related medical imaging concerns, summarized reviews on deep learning application to COVID-19
620 medical imaging analysis, and listed and described contributions to this domain. Finally, they discussed the challenges
621 experienced in the research domain and made suggestions for future research.

622

623 **6. Survey and bibliometric analysis**

624 **6.1 Survey analysis**

625 In this survey, we review the projects that used machine learning to fight COVID-19 from a different perspective. We
626 only considered published papers in reputable journals, and conferences, and no preprint papers uploaded in preprint
627 server was used in the survey. We apprised 30 studies that reported the description of the machine learning approach
628 to fighting COVID-19. We found that machine learning has made an inroad into fighting COVID-19 from a different
629 aspect with potential for real-life applications to curtail the negative effect of COVID-19. Machine learning algorithms
630 such as CNN, LSTM, and ANN that are utilized in fighting COVID-19 mostly reported excellent performance
631 compared with the baseline approaches. Many of the studies complained about the scarcity of sufficient data to carry
632 out large-scale study because of the novelty of the COVID-19 pandemic.

633 We found that various studies used different COVID-19 data. Figure 4 depicts the type of data used in different studies
634 that applied machine learning algorithm to develop different models for fighting COVID-19 pandemic. The data used
635 to plot Figure 4 were extracted from machine learning research on COVID-19 (refer to Table 6). The longest bars
636 show that X-rays and CT scans have the highest patronage from the studies. Many of the studies used deep learning
637 algorithms, e.g., CNN and LSTM, for the diagnosis of COVID-19 on X-rays and CT scans. The evaluation indicated
638 the excellent performance of the algorithms in detecting COVID-19 on X-rays and CT scan images.

639

640   **Figure 4: Visual representation of COVID-19 data extracted from different projects**

641 The CT scan has a great value in the screening, diagnosis, and follow up of patients with COVID-19. The CT scan
642 has now been added as a criterion for diagnosing of COVID-19 (Liu et al., 2020; Li et al., 2020). The X-rays with
643 COVID-19 pandemic data project by Cohen hosted on GitHub is receiving unprecedented attention from the research
644 community for accessing freely available data. Figure 5 presents the frequency of machine learning algorithms adopted
645 to fight COVID-19. The longest bar indicates that CNN received the most considerable attention from the researchers
646 working in this domain to fight COVID-19. The likely reason why CNN has the highest number of applications is that
647 most of the data used in detecting COVID-19 infection in patients are images (see Figure 4). CNN is well known for
648 its robustness, effectiveness, and efficiency in image processing compared with other conventional machine learning
649 algorithms because of its automated feature engineering and high performance. The CNN variant suitable for the
650 diagnosis of COVID-19 from X-ray and CT scan images is ResNet. However, many of the studies did not provide the
651 specific type of CNN adopted for the diagnosis of the COVID-19 from X-ray and CT scan images (see Table 3).

652

653 **Figure 5: Machine learning algorithms adopted in fighting COVID-19**

654

655 Figure 6 shows the different aspects where machine learning algorithms were applied in fighting COVID-19. We
656 found that the studies mainly adopted machine learning algorithms in developing COVID-19 diagnosis tools, decision
657 support system, drug development, and detection from protein sequence. The most extended portion of the pie chart
658 indicates that diagnostic tools attracted the most considerable attention, showing the quest for diagnostic tools in the
659 fight against COVID-19 pandemic because the match starts with a diagnosis before the appropriate treatment is
660 administered to save a life, and incorrect diagnosis can lead to inappropriate medication, resulting in further health
661 complications. Most of the studies that adopted machine learning to develop diagnostic tools intended to reduce the
662 workload of radiologists, improve the speed of diagnosis, automate the COVID-19 diagnostic process, reduce the cost
663 compared with traditional laboratory tests, and help healthcare workers in making critical decisions.

664

665 **Figure 6: Different aspects of machine learning applications in fighting the COVID-19 pandemic**

666

667 The studies argued that the diagnostic tool could reduce the exposure of healthcare workers to patients with COVID-
668 19, thus decreasing the risk of spreading COVID-19 to healthcare workers. The second part of the pie chart with the
669 most substantial portion is the decision support system for detecting the rate of spread of the virus, confirmed cases,
670 mortalities, and recovered cases. This information from the decision support system can help the government
671 functionaries, policymakers, decision makers, and other stakeholders in formulating policy that can help fight COVID-
672 19 pandemic.

673

674 **Figure 7: Trend of publications on machine learning applications in fighting COVID-19**

675

676 Figure 7 shows that the publications on the adoption of machine learning to fight COVID-19 started appearing in 2020
677 with no publications in 2019, although COVID-19 started appearing in late 2019. This situation is likely because a
678 new virus-like the COVID-19 pandemic at the initial stage lacks data and has scant information and uncertainties.

679 **6.2 Bibliometric analysis**

680 The primary purpose of conducting a bibliometric analysis study in this study is to reflect the trend of rapidly emerging
681 topics on COVID-19 research, where substantial research activity has already begun extensively during the early stage
682 of the outbreak. Another significance of the bibliometric analysis method presented is to aid in the mapping of research
683 situation on coronavirus disease as reported in several scientific works of literature by the research community. In this
684 section, we present the bibliographic coupling among different article items on machine learning to fight COVID-19.
685 The link between the items on the constructed map corresponds to the weight between them either in terms of the
686 number of publications, common references, or co-citations. These items may belong to a group or a cluster. In the
687 visualization, items within the same cluster are marked with the same color, and colors indicate the cluster to which a
688 journal was assigned by the clustering technique implemented by the VOSviewer software. The circular node may
689 represent the items, and their sizes may vary depending on the weight of the article.

690 *6.2.1 Prolific authors*

691 The bibliographic coupling between the top 25 authors is shown in Figure 8. The two clusters, namely, red and green,
692 correspond to all authors working on similar research fields "COVID-19" and citing the same source in their reference
693 listings. The similarity in cluster color for the authors also implies that the degree of overlap between the reference

694   lists of publications of these authors is higher. Figure 8 shows the visible names, and other names may not be
695   represented in the constructed map.

696

697   **Figure 8: Bibliographic coupling among authors**

698

699   *6.2.2 Productive countries*

700   Figure 9 shows the bibliographic coupling of the topmost productive countries. Here, bibliographic coupling indicates
701   that a common reference list in the papers published by these countries. The five clusters are represented by six colors.
702   Red represents China and the USA with the highest strength in terms of contributions, after which comes India and
703   Iran as the next countries within the red node. Green represents Hong Kong, which appears to have the highest strength,
704   whereas blue is for the United Kingdom and Saudi Arabia that have the highest strength. Yellow denotes Japan,
705   Singapore, Thailand, and Taiwan as the highest contributors. Purple refers to Italy and Canada as the two contributing
706   countries. The link between the red and green clusters are thicker compared with that between the blue and red clusters,
707   or between the blue and purple clusters. The thickness of the link simply depicts the degree of intersection of the
708   literature work between the different locations or countries.
709

710
711
712   **Figure 9: Bibliographic coupling among countries**
713
714

715   *6.2.3 Collaborating institution*

716   Figure 10 shows the bibliographic coupling of the network of collaborating institutions that are affiliated with at least
717   four documents on COVID-19 research. Map construction and analysis show that of the 949 institutions, 28 meet the
718   thresholds of collaborating institutions, including the Guangzhou Center for Disease Control (or Chinese Center for
719   Disease Control and Prevention), Huazhong University of Science and Technology, Wuhan University, Capital
720   Medical University, Chinese Academy of Medical Sciences, University of Hong Kong, Sun Yat-Sen University, and
721   Fudan University. Two clusters are identified by red and green colors. Institutions that fall under the same groups
722   appear to have similar literature background or worked on related research fields.
723

724

725   **Figure 10: Bibliographic coupling among institutions**

726   *6.2.4 Journals*

727   Bibliographic coupling between journals implies that the papers published in these journals have more common
728   reference lists. Three clusters are depicted on the map with red, blue, and green colors. The links with the highest
729   strength occur between Emerging Microbes Journal, Journal of Virology, and Journal of Infection. This link is closely
730   followed by the links between Eurosurveillance and Journal of Infection, Archive of Academic Emergency Medicine,
731   Chinese Medical Journal, and The Lancet. The Journal of Infection Control and Hospital and Journal of Hospital
732   Infection form the weakest networks of a cluster. Figure 11 illustrates the bibliographic coupling between the
733   considered journals.

734

735   **Figure 11: Bibliographic coupling among journals**

736

737
738 *6.2.5 Co-authorship and authors*
739
740 Figure 12 illustrates the co-authors and author map visualization. This analysis aims to produce the visualization of
741 all the major authors publishing together or working on similar research fields. The analysis type is co-authorship, and
742 the unit of analysis is authors. The threshold of the minimum number of papers by an author is 25. Network
743 construction and analysis shows that of 2,381 authors, only 9 authors meet the limits. However, the most extensive set
744 of connected entities consists of only 8 authors, whose visual representation is depicted in Figure 12, where only one
745 cluster is denoted by red color. The connected link illustrates that these authors have collaborated on the same project
746 or worked on the same research with a similar focus. The thickness of the link between these three authors indicates
747 more common publications.
748
749
750 **Figure 12: Co-authorship and authors' analysis**
751
752
753 *6.2.6 Impact analysis*
754 Figure 13 illustrates the citation analysis among authors' institutions. Six clusters are represented using different
755 colors. The red cluster has the highest number of author citations from two institutions, namely, the Huazhong
756 University of Science and Technology, Wuhan University (*State Key Laboratory of Virology*), and the Department of
757 Microbiology, University of Hong Kong. Figure 14 shows the bibliometric analyses of author citations by journal
758 sources. A link between two journal sources indicates the citation connectivity between the two sources.
759
760 **Figure 13: Author citation by institution**
761
762 **Figure 14: Author citation by journal source**
763
764
765 The connected links between the Journal of Virology and the New England Journal of Medicine in Figure 14 reveal
766 that a publication from the Journal of Virology has cited another publication that is published in the New England
767 Journal of Medicine or vice versa. The thickness and link strength signify more numbers of citation among the clusters.
768 Therefore, among the different clusters identified in the analysis, the Journal of Virology is the top-cited source by
769 publication from other journal sources.

770 **7. Challenges and future research opportunities**

771 In this section, we present challenges and future research prospects. More so, Figure 15 describes the course of
772 conducting the literature survey and opportunities for future research with the possibility of solving the challenges to
773 help expert researchers easily identify areas that need development. The challenges and future research opportunities
774 are presented as follows:

775 ***Lack of sufficient COVID-19 data****:* The primary concern with the research in COVID-19 is the barrier prompted by
776 the lack of adequate COVID-19 clinical data (Alimadadi et al., 2020; Mei et al., 2020; Ayyoubzadeh et al., 2020; Fong
777 et al., 2020; Oh et al., 2020; Togaçar et al., 2020; Ucara and Korkmazb, 2020; Belfiore et al., 2020). However, an in-
778 depth analysis of patients with COVID-19 requires much more data (Apostolopoulos and Mpesiana, 2020). Data is
779 the key component in machine learning. Machine learning approaches typically experience a limitation in their
780 efficiency and effectiveness in solving machine learning problems without sufficient data. Therefore, insufficient
781 COVID-19 clinical data can limit the performance of specific machine learning algorithms, such as deep learning
782 algorithms that require large-scale data. In this case, developing machine-learning-based COVID-19 diagnostic and
783 prognosis tools, and therapeutic approaches to curtail COVID-19, and predicting a future pandemic can face a severe

784 challenge in terms of performance due to insufficient COVID-19 clinical data. Alimadadi et al. (2020) suggested
785 global collaborations among stakeholders to build COVID-19 clinical database and mitigate the issue of inadequate
786 COVID-19 clinical data. Existing biobanks containing the data of patients with COVID-19 are integrated with
787 COVID-19 clinical data. We suggest that researchers use GAN to generate additional X-rays and CT scan images for
788 COVID-19 to obtain sufficient data for building COVID-19 diagnosis tools. For example, Loey et al. (2020) were
789 motivated by insufficient data and used GAN to generate more X-ray images and develop a COVID-19 diagnostic
790 tool.

791 ***Visualization from X-rays and CT Scan***: Figure 4 shows that X-ray and CT scan are the two primary clinical data for
792 detecting COVID-19 infection in patients. Distinguishing patients with COVID-19 and mild symptoms from
793 pneumonia on X-ray images could be visualized inaccurately or cannot be visualized totally (Apostolopoulos and
794 Mpesiana, 2020). We suggest that researchers propose machine learning strategies that can accurately differentiate
795 patients with COVID-19 and mild symptoms from patients with pneumonia symptoms based on X-ray images.
796 COVID-19 that is caused by coronavirus might have a CT scan image characteristic similar to other pneumonia caused
797 by a different virus. In the future, the performance of CNN should be evaluated in classifying COVID-19 and viral
798 pneumonia with RT–PCR (Li et al., 2020b)

799 ***Uncertainties***: When a new pandemic breaks out, it comes with limited information and very high uncertainly, unlike
800 the commonly known influenza. Therefore, knowledge regarding the new epidemic is not sufficient due to the absence
801 of a prior case that is the same as the recent pandemic. In the case of COVID-19, many of the decision makers relied
802 on SARS for reference because of the similarity, even though it is considerably different from COVID-19. The new
803 pandemic typically poses a challenge to data analytics, considering its limited information and geographical and
804 temporal evolving of the recent epidemic. Therefore, an accurate model for predicting the future behavior of a
805 pandemic becomes challenging due to uncertainty (Fong et al., 2020). We suggest that researchers propose a new
806 pandemic forecasting model based on active learning in machine learning to reduce the level of uncertainty, typically
807 accompanying new pandemics such as COVID-19.

808 ***Non-uniform pattern***: Liu et al. (2020) applied Susceptible, Exposed, Infectious, Recovered (SEIR) for modeling
809 COVID-19. However, the SEIR model could not capture the complete number of infected cases, while the study
810 ignored imported COVID-19 confirmed cases. SEIR was based on the people's natural distribution and cannot apply
811 to welfare institute an example of different population distribution. The epidemiological trend of COVID-19 was not
812 predicted accurately by the SEIR model under the viral mutation and specific ant-viral therapy development scenario.
813 The SEIR model was unable to simulate non-uniform patterns, such as the issue of increasing medical professionals
814 and bed capacity (Liu et al., 2020). We suggest that researchers propose a machine-learning-based strategy for
815 handling the non-uniform pattern in the future and consider all the other factors not considered in the study.

816 ***Insufficient regional data***: Adequate COVID-19 data for a particular region are lacking because the capacity to gather
817 reliable data is not uniform across regions worldwide. This situation can bring a challenge to the region without
818 available COVID-19 data. We suggest that researchers apply the cross-population train–test model because a model
819 trained in a different region can be used to detect COVID-19 in a different region. For example, the model trained to
820 detect the new virus in Wuhan, China, can be used in Italy (Santosh, 2020).

821

822   **Figure 15: Visual representation of challenges**

823

824 ***Image resolution***: The resolution of the X-ray images affects the performance of the machine learning algorithm.
825 Dealing with low-resolution images typically poses a challenge to the machine learning approach. Variable size of the
826 resolution dimension has a negative effect. Successful performance cannot be achieved if the input images of the data
827 have different sizes. The original image resolution dimension, structured images, and stacking technique need to be
828 the same (Togaçar et al., 2020). We suggest high-resolution X-ray images for developing COVID-19 diagnostic and
829 prognosis system with the ability to work with low-resolution X-ray images.

830   ***Outliers and noise***: At the early phase of COVID-19, the COVID-19 data contained many outliers and much noise
831   (Tuli et al., 2020). An outlier in data is a subset of the data that appears with inconsistencies from the remaining data.
832   Outliers typically lower the fit of a statistical model (Bellazzi et al., 1998). The presence of outliers and noise in
833   COVID-19 makes predicting the correct number of COVID-19 cases challenging (Tuli et al., 2020). Dealing with
834   outliers and noise in data increases data engineering efforts and expenses. We suggest that researchers propose a robust
835   machine learning approach that can effectively handle outliers and noise in COVID-19 data.

836   ***Transparency and interpretability***: The limitation of deep learning algorithms is a deficiency in terms of transparency
837   and interpretability. For instance, knowing the image features that are applied to decide the output of the deep learning
838   algorithms is not possible. The unique features used by the deep learning algorithm to differentiate COVID-19 from
839   CAP cannot be sufficiently visualized by the heatmap, although the heatmap is used to visualize region in images that
840   led to the algorithm output (Li et al., 2020b). Images, especially X-rays and CT scans, are heavily relied on in detecting
841   COVID-19. We suggest that researchers propose explainable deep learning algorithms for the detection of COVID-
842   19 to instill transparency and interpretation in deep learning algorithms.

843   ***Misdiagnosis***: The application of a deep learning algorithm to detect COVID-19 on a chest CT scan has the possibility
844   of misdiagnosis (Wu et al., 2020) because of the similarity of the COVID-19 symptoms with other types of pneumonia
845   (Belfiore et al., 2020). Incorrect diagnoses can mislead the health professional in deciding and lead to inappropriate
846   medication, further complicating the health condition of the patient with COVID-19. We suggest that researchers
847   combine the CT scan diagnosis using deep learning algorithm with clinical information such as the nucleic acid
848   detection results, clinical symptoms, epidemiology, and laboratory indicators to avoid misdiagnosis (Wu et al., 2020).

849   ***Resource allocation***: Resource allocation is a challenge as the COVID-19 pandemic keeps spreading because the
850   increase in the number of patients means more resources are required to take care of them. The allocation of limited
851   resources in a rapidly expanding pandemic entails a difficult decision for the distribution of scarce resources (Jiang et
852   al., 2020). The epicenters of the COVID-19 are challenged with resource problems of shortage of beds, gowns, masks,
853   medical staff, and ventilators (Ahuja et al., 2020). We propose the development of a machine learning decision support
854   system to help in crucial decisions on resource allocation.

855   **8. Conclusions**

856   In this study, we propose a survey, including a bibliometric analysis of the adoption of machine learning, to fight
857   COVID-19. The concise summary of the projects that adopted machine learning to fight COVID-19, sources of
858   COVID-19 datasets, new comprehensive taxonomy, synthesis and analysis, and bibliometric analysis is presented.
859   The results reveal that COVID-19 diagnostic tools received the most considerable attention from researchers, and
860   energy and resources are more dispensed toward automated COVID-19 diagnostic tools. By contrast, COVID-19
861   drugs and vaccine development remain grossly underexploited. The algorithm predominantly utilized by the
862   researchers in developing the diagnostic tool is CNN mainly from X-rays and CT scan images. The most suitable
863   CNN architecture for the detection of COVID-19 from the X-ray and CT scan images is ResNet. The challenges
864   hindering practical work on machine learning to fight COVID-19 and a new perspective to solve the identified
865   problems are presented in the study. We believe that our survey with bibliometric analysis could enable researchers
866   to determine areas that need further development and identify potential collaborators at author, country, and
867   institutional levels.

868   Based on the bibliometric analysis conducted on the global scientific research output on COVID-19 disease spread
869   and preventive measures, the analysis results reveal that most of the research outputs were published in prestigious
870   journals with high influence factors. These journals include The Lancet, Journal of Medical Virology, and
871   Eurosurveillance. The bibliometric analysis also shows the focused subjects in various aspects of COVID-19 infection
872   transmission, diagnosis, treatment, prevention, and its complications. Other prominent features include strong
873   collaboration among research institutions, universities, and co-authorships among researchers across the globe.

874   Machine learning algorithms have many practical applications in medicine, and novel contributions from different
875   researchers are still evolving and growing exponentially in a bid to satisfy the essential clinical needs of the individual
876   patients, as it is the case with its application to fighting the COVID-19 pandemic. As a way forward, we suggest an

877 in-depth machine learning application review that would focus on the critical analysis of the novel coronavirus disease
878 and other related cases of global pandemics.

879 **Declaration of Competing Interest**

880 The authors declare that they have no known competing financial interests or personal relationships that could have
881 appeared to influence the work reported in this paper.

882

883 **References**
884 Ahuja, A. S., Reddy, V. P., & Marques, O. (2020). Artificial Intelligence and COVID-19: A Multidisciplinary
885   Approach. Integrative medical research. https://doi.org/10.1016/j.imr.2020.100434
886 Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., ... & Xia, L. (2020). Correlation of chest CT and RT-PCR
887   testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. *Radiology*, 200642.
888 Alimadadi, A., Aryal, S., Manandhar, I., Munroe, P. B., Joe, B., & Cheng, X. (2020). Artificial intelligence and
889   machine learning to fight COVID-19. *Physiol Genomics*, *52: 200 –202*.
890 Ardakani, A. A., Kanafi, A. R., Acharya, U. R., Khadem, N., & Mohammadi, A. (2020). Application of deep learning
891   technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional
892   neural networks. *Computers in Biology and Medicine*, 103795.
893 Asif, S., & Wenhui, Y. (2020). Automatic Detection of COVID-19 Using X-ray Images with Deep Convolutional
894   Neural Networks and Machine Learning. Computers in Biology and Medicine,Volume 121, June 2020,
895   103792.
896 Ayyoubzadeh, S. M., Ayyoubzadeh, S. M., Zahedi, H., Ahmadi, M., & Kalhori, S. R. N. (2020). Predicting COVID-
897   19 incidence through analysis of google trends data in iran: data mining and deep learning pilot study. *JMIR
898   Public Health and Surveillance*, *6*(2), e18828.
899 Ayyoubzadeh, S. M., Ayyoubzadeh, S. M., Zahedi, H., Ahmadi, M., & Kalhori, S. R. N. (2020). Predicting COVID-
900   19 incidence through analysis of google trends data in iran: data mining and deep learning pilot study. JMIR
901   Public Health and Surveillance, 6(2), e18828.
902 Barbosa, R. D. M., & Fernandes, M. A. (2020). Chaos game representation dataset of SARS-CoV-2 genome. *Data in
903   Brief*, 105618.
904 Basheer, I. A., & Hajmeer, M. (2000). *Artificial neural networks : fundamentals , computing , design , and application*.
905   *43*(1), 3–31.
906 Belfiore, M. P., Urraro, F., Grassi, R., Giacobbe, G., Patelli, G., Cappabianca, S., & Reginelli, A. (2020). Artificial
907   intelligence to codify lung CT in Covid-19 patients. *La radiologia medica*, 1-5.
908 Bellazzi, R., Guglielmann, R., & Ironi, L. (1998). Qualitative and Fuzzy Reasoning for identifying non-linear
909   physiological systems: an application to intracellular thiamine kinetics. In *5 th International Workshop on
910   Intelligent Data Analysis in Medicine and Pharmacology*.
911 Bernheim, A., Mei, X., Huang, M., Yang, Y., Fayad, Z. A., Zhang, N., ... & Li, S. (2020). Chest CT findings in
912   coronavirus disease-19 (COVID-19): relationship to duration of infection. *Radiology*, 200463.
913 Butt, C., Gill, J., Chun, D., & Babu, B. A. (2020). Deep learning system to screen coronavirus disease 2019 pneumonia.
914   *Applied Intelligence*, 1.
915 Casey, B. (2020). How good is radiography for COVID-19 detection? AuntMinnie.com.
916 Chahrour, M., Assi, S., Bejjani, M., Nasrallah, A. A., Salhab, H., Fares, M., & Khachfe, H. H. (2020). A bibliometric
917   analysis of Covid-19 research activity: A call for increased output. *Cureus*, *12*(3).
918 Chimmula, V. K. R., & Zhang, L. (2020). Time series forecasting of covid-19 transmission in canada using lstm
919   networks. *Chaos, Solitons & Fractals*, 109864.
920 Corman, V. M., Landt, O., Kaiser, M., Molenkamp, R., Meijer, A., Chu, D. K., ... & Mulders, D. G. (2020). Detection
921   of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro surveillance*, *25*(3), 2000045.
922 Dananjayan, S., & Raj, G. M. (2020). Artificial Intelligence during a pandemic: The COVID-19 example. *The
923   International Journal of Health Planning and Management*.
924 Das, K. M., Lee, E. Y., Singh, R., Enani, M. A., Al Dossari, K., Van Gorkom, K., ... & Langer, R. D. (2017). Follow-up
925   chest radiographic findings in patients with MERS-CoV after recovery. *The Indian journal of radiology &
926   imaging*, *27*(3), 342.
927 Deng, L. (2014). A tutorial survey of architectures, algorithms, and applications for deep learning. *APSIPA
928   Transactions on Signal and Information Processing*, *3*.

929 Dong, D., Tang, Z., Wang, S., Hui, H., Gong, L., Lu, Y., ... & Jin, R. (2020). The role of imaging in the detection and
930   management of COVID-19: a review. *IEEE reviews in biomedical engineering*.
931 Duan, Y. N., & Qin, J. (2020). Pre-and posttreatment chest CT findings: 2019 novel coronavirus (2019-nCoV)
932   pneumonia. *Radiology*, *295*(1), 21-21.
933 Elavarasan, R. M., & Pugazhendhi, R. (2020). Restructured society and environment: A review on potential
934   technological strategies to control the COVID-19 pandemic. *Science of The Total Environment*, 138858.
935 Fang, Y., Zhang, H., Xie, J., Lin, M., Ying, L., Pang, P., & Ji, W. (2020). Sensitivity of chest CT for COVID-19:
936   comparison to RT-PCR. *Radiology*, 200432.
937 Farhat, H., Sakr, G. E., & Kilany, R. (2020). Deep learning applications in pulmonary medical imaging: recent updates
938   and insights on COVID-19. Machine Vision and Applications, 31(6), 1-42
939 Fong, S. J., Li, G., Dey, N., Crespo, R. G., & Herrera-Viedma, E. (2020). Composite Monte Carlo decision making
940   under high uncertainty of novel coronavirus epidemic using hybridized deep learning and fuzzy rule
941   induction. *Applied Soft Computing*, 106282.
942 Gonzalez-Dias, P., Lee, E. K., Sorgi, S., de Lima, D. S., Urbanski, A. H., Silveira, E. L., & Nakaya, H. I. (2020).
943   Methods for predicting vaccine immunogenicity and reactogenicity. *Human Vaccines &*
944   *Immunotherapeutics*, *16*(2), 269-276.
945 Gorbalenya, A. E. (2020). Severe acute respiratory syndrome-related coronavirus–The species and its viruses, a
946   statement of the Coronavirus Study Group. *BioRxiv*.
947 Greff, K., Srivastava, R. K., Koutn, J., Steunebrink, B. R., & Schmidhuber, J. (2016). LSTM : A Search Space
948   Odyssey. *IEEE Transactions on Neural Networks and Learning Systems, 28(10), 2222-2232.*
949 Guan, W. J., Ni, Z. Y., Hu, Y., Liang, W. H., Ou, C. Q., He, J. X., ... & Du, B. (2020). Clinical characteristics of
950   coronavirus disease 2019 in China. *New England journal of medicine*, *382*(18), 1708-1720.
951 Haque, I. R. I., & Neubert, J. (2020). Deep learning approaches to biomedical image segmentation. *Informatics in*
952   *Medicine Unlocked*, *18*, 100297.
953 Holshue, M. L., DeBolt, C., Lindquist, S., Lofy, K. H., Wiesman, J., Bruce, H., . . . Tural, A. (2020). First case of
954   2019 novel coronavirus in the United States. *New England Journal of Medicine*.
955 Hossain, M. M. (2020). Current status of global research on novel coronavirus disease (Covid-19): A bibliometric
956   analysis and knowledge mapping. *Hossain MM. Current status of global research on novel coronavirus disease*
957   *(COVID-19): a bibliometric analysis and knowledge mapping version 1.*
958 Hu, S., Gao, Y., Niu, Z., Jiang, Y., Li, L., Xiao, X., ... & Ye, H. (2020). Weakly supervised deep learning for covid-
959   19 infection detection and classification from ct images. IEEE Access, 8, 118869-118883.
960 Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., ... & Cheng, Z. (2020b). Clinical features of patients infected
961   with 2019 novel coronavirus in Wuhan, China. *The lancet*, *395*(10223), 497-506.
962 Huang, L., Han, R., Ai, T., Yu, P., Kang, H., Tao, Q., & Xia, L. (2020). Serial quantitative chest ct assessment of
963   covid-19: Deep-learning approach. *Radiology: Cardiothoracic Imaging*, *2*(2), e200075.
964 Hui, D. S., Azhar, E. I., Madani, T. A., Ntoumi, F., Kock, R., Dar, O., ... & Zumla, A. (2020). The continuing 2019-
965   nCoV epidemic threat of novel coronaviruses to global health—The latest 2019 novel coronavirus outbreak in
966   Wuhan, China. *International Journal of Infectious Diseases*, *91*, 264-266.
967 Hurt, B., Kligerman, S., & Hsiao, A. (2020). Deep Learning Localization of Pneumonia: 2019 Coronavirus (COVID-
968   19) Outbreak. *Journal of Thoracic Imaging*, *35*(3), W87-W89.
969 Imai, N., Dorigatti, I., Cori, A., Donnelly, C., Riley, S., & Ferguson, N. (2020). Report 2: Estimating the potential
970   total number of novel Coronavirus cases in Wuhan City, China.
971 Inui, S., Fujikawa, A., Jitsu, M., Kunishima, N., Watanabe, S., Suzuki, Y., ... & Uwabe, Y. (2020). Chest CT findings
972   in cases from the cruise ship "Diamond Princess" with coronavirus disease 2019 (COVID-19). *Radiology:*
973   *Cardiothoracic Imaging*, *2*(2), e200110.
974 Jaiswal, A., Gianchandani, N., Singh, D., Kumar, V., & Kaur, M. (2020). Classification of the COVID-19 infected
975   patients using DenseNet201 based deep transfer learning. Journal of Biomolecular Structure and Dynamics, 1-
976   8.
977 Jiang, X., Coffee, M., Bari, A., Wang, J., Jiang, X., Shi, J., ... & He, G. (2020). Towards an artificial intelligence
978   framework for data-driven prediction of coronavirus clinical severity. *Computers, Materials & Continua*,
979   *63*(1), 537-551.
980 Karim, F., Majumdar, S., Darabi, H., & Chen, S. (2018). LSTM Fully Convolutional Networks for Time Series
981   Classification. *IEEE Access*, *6*, 1662–1669.
982 Ke, Y. Y., Peng, T. T., Yeh, T. K., Huang, W. Z., Chang, S. E., Wu, S. H., ... & Lin, W. H. (2020). Artificial
983   intelligence approach fighting COVID-19 with repurposing drugs. *Biomedical Journal*.

984 Klein, S. L., Jedlicka, A., & Pekosz, A. (2010). The Xs and Y of immune responses to viral vaccines. *The Lancet*
985   *infectious diseases*, *10*(5), 338-349.
986 Kooraki, S., Hosseiny, M., Myers, L., & Gholamrezanezhad, A. (2020). Coronavirus (COVID-19) outbreak: what the
987   department of radiology should know. *Journal of the American college of radiology*.
988 Kumar, A., Gupta, P. K., & Srivastava, A. (2020). A review of modern technologies for tackling COVID-19 pandemic.
989   *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*.
990 Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.
991 Li, D., Wang, D., Dong, J., Wang, N., Huang, H., Xu, H., & Xia, C. (2020). False-negative results of real-time reverse-
992   transcriptase polymerase chain reaction for severe acute respiratory syndrome coronavirus 2: role of deep-
993   learning-based CT diagnosis and insights from two cases. *Korean journal of radiology*, *21*(4), 505-508.
994 Li, L., Qin, L., Xu, Z., Yin, Y., Wang, X., Kong, B., ... & Cao, K. (2020b). Artificial intelligence distinguishes COVID-
995   19 from community acquired pneumonia on chest CT. *Radiology*, 200905.
996 Li, Z., Zhong, Z., Li, Y., Zhang, T., Gao, L., Jin, D., ... & Xiao, J. (2020). From Community Acquired Pneumonia to
997   COVID-19: A Deep Learning Based Method for Quantitative Analysis of COVID-19 on thick-section CT
998   Scans. Eur Radiol (2020). https://doi.org/10.1007/s00330-020-07042-x.
999 Liu, F., Zhang, Q., Huang, C., Shi, C., Wang, L., Shi, N., ... & Song, F. (2020). CT quantification of pneumonia lesions
1000   in early days predicts progression to severe illness in a cohort of COVID-19 patients. *Theranostics*, *10*(12),
1001   5613.
1002 Liu, G., Xu, Y. A. N., He, Z., Rao, Y., Xia, J., & Fan, L. (2019). Deep Learning-Based Channel Prediction for Edge
1003   Computing Networks Toward Intelligent Connected Vehicles. *IEEE Access*, *7*, 114487–114495.
1004 Liu, J., Yu, H., & Zhang, S. (2020). The indispensable role of chest CT in the detection of coronavirus disease 2019
1005   (COVID-19). European journal of nuclear medicine and molecular imaging.
1006 Liu, Z., Huang, S., Lu, W., Su, Z., Yin, X., Liang, H., & Zhang, H. (2020b). Modeling the trend of coronavirus disease
1007   2019 and restoration of operational capability of metropolitan medical service in China: a machine learning
1008   and mathematical model-based analysis. *Global Health Research and Policy*, *5*, 1-11.
1009 Livingstone, D. J. (Ed. . (2008). *Artificial neural networks: methods and applications*. Totowa: NJ:Humana Press.
1010 Loey, M., Smarandache, F., & M Khalifa, N. E. (2020). Within the Lack of Chest COVID-19 X-ray Dataset: A Novel
1011   Detection Model Based on GAN and Deep Transfer Learning. *Symmetry*, *12*(4), 651.
1012 Lou, J., Tian, S. J., Niu, S. M., Kang, X. Q., Lian, H. X., Zhang, L. X., & Zhang, J. J. (2020). Coronavirus disease
1013   2019: a bibliometric analysis and review. *Eur Rev Med Pharmacol Sci*, *24*(6), 3411-21.
1014 Mei, X., Lee, H. C., Diao, K. Y., Huang, M., Lin, B., Liu, C., ... & Bernheim, A. (2020). Artificial intelligence–enabled
1015   rapid diagnosis of patients with COVID-19. *Nature Medicine*, 1-5.
1016 Mohammadi, M., Al-Fuqaha, A., Sorour, S., & Guizani, M. (2018). Deep learning for IoT big data and streaming
1017   analytics: A survey. *IEEE Communications Surveys & Tutorials*, *20*(4), 2923-2960.
1018 Oh, Y., Park, S., & Ye, J. C. (2020). Deep learning covid-19 features on cxr using limited training data sets. *IEEE*
1019   *Transactions on Medical Imaging*.
1020 Ozturk, T., Talo, M., Yildirim, E. A., Baloglu, U. B., Yildirim, O., & Acharya, U. R. (2020). Automated detection of
1021   COVID-19 cases using deep neural networks with X-ray images. *Computers in Biology and Medicine*, 103792.
1022 Pan, F., Ye, T., Sun, P., Gui, S., Liang, B., Li, L., ... & Zheng, C. (2020). Time course of lung changes on chest CT
1023   during recovery from 2019 novel coronavirus (COVID-19) pneumonia. *Radiology*, 200370.
1024 Park, Y., Casey, D., Joshi, I., Zhu, J., & Cheng, F. (2020). Emergence of new disease–how can artificial intelligence
1025   help?. *Trends in Molecular Medicine*.
1026 Patel, A., & Jernigan, D. B. (2020). Initial public health response and interim clinical guidance for the 2019 novel
1027   coronavirus outbreak—United States, December 31, 2019–February 4, 2020. *Morbidity and Mortality Weekly*
1028   *Report, 69*(5), 140.
1029 Peng, Q. Y., Wang, X. T., Zhang, L. N., & Chinese Critical Care Ultrasound Study Group. (2020). Findings of lung
1030   ultrasonography of novel corona virus pneumonia during the 2019–2020 epidemic. *Intensive care medicine*, 1.
1031 Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M. P., … Iyengar, S. S. (2018). A Survey on Deep
1032   Learning : Algorithms , Techniques , and Applications. *ACM Computing Surveys (CSUR)*, *51*(5), 1–36.
1033 Qiang, X. L., Xu, P., Fang, G., Liu, W. B., & Kou, Z. (2020). Using the spike protein feature to predict infection risk
1034   and monitor the evolutionary dynamic of coronavirus. *Infectious Diseases of Poverty*, *9*(1), 1-8.
1035 Rahaman, M. M., Li, C., Yao, Y., Kulwa, F., Rahman, M. A., Wang, Q., ... & Zhao, X. (2020). Identification of
1036   COVID-19 samples from chest X-Ray images using deep learning: A comparison of transfer learning
1037   approaches. *Journal of X-Ray Science and Technology*, IOS Press, 1-19.

1038 Randhawa, G. S., Soltysiak, M. P., El Roz, H., de Souza, C. P., Hill, K. A., & Kari, L. (2020). Machine learning using
1039   intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study. *Plos one*, *15*(4),
1040   e0232391.
1041 Randhawa, G. S., Soltysiak, M. P., El Roz, H., de Souza, C. P., Hill, K. A., & Kari, L. (2020). Machine learning using
1042   intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study. Plos one, 15(4),
1043   e0232391.
1044 Rao, A. S. S., & Vazquez, J. A. (2020). Identification of COVID-19 can be quicker through artificial intelligence
1045   framework using a mobile phone–based survey when cities and towns are under quarantine. *Infection Control*
1046   *& Hospital Epidemiology*, 1-5.
1047 Read, J. M., Bridgen, J. R., Cummings, D. A., Ho, A., & Jewell, C. P. (2020). Novel coronavirus 2019-nCoV: early
1048   estimation of epidemiological parameters and epidemic predictions. *medRxiv*.
1049 Ribeiro, M. H. D. M., da Silva, R. G., Mariani, V. C., & dos Santos Coelho, L. (2020). Short-term forecasting COVID-
1050   19 cumulative confirmed cases: Perspectives for Brazil. *Chaos, Solitons & Fractals*, 109853.
1051 Rodriguez-Morales, A. J., Cardona-Ospina, J. A., Gutiérrez-Ocampo, E., Villamizar-Peña, R., Holguin-Rivera, Y.,
1052   Escalera-Antezana, J. P., ... & Paniz-Mondolfi, A. (2020). Clinical, laboratory and imaging features of COVID-
1053   19: A systematic review and meta-analysis. *Travel medicine and infectious disease*, 101623.
1054 Sak, H., Senior, A., & Beaufays, F. (2014). Long Short-Term Memory Recurrent Neural Network Architectures for
1055   Large Scale Acoustic Modeling Has. *Fifteenth Annual Conference of the International Speech Communication*
1056   *Association*.
1057 Salman, F. M., Abu-Naser, S. S., Alajrami, E., Abu-Nasser, B. S., & Alashqar, B. A. (2020). Covid-19 detection using
1058   artificial intelligence.
1059 Santosh, K. C. (2020). AI-driven tools for coronavirus outbreak: need of active learning and cross-population train/test
1060   models on multitudinal/multimodal data. *Journal of medical systems*, *44*(5), 1-5.
1061 Shi, F., Wang, J., Shi, J., Wu, Z., Wang, Q., Tang, Z., ... & Shen, D. (2020). Review of artificial intelligence techniques
1062   in imaging data acquisition, segmentation and diagnosis for covid-19. *IEEE reviews in biomedical engineering*.
1063 Singh, D., Kumar, V., & Kaur, M. (2020). Classification of COVID-19 patients from chest CT images using multi-
1064   objective differential evolution–based convolutional neural networks. *European Journal of Clinical*
1065   *Microbiology & Infectious Diseases*, 1-11.
1066 Singh, D., Kumar, V., & Kaur, M. (2020). Classification of COVID-19 patients from chest CT images using multi-
1067   objective differential evolution–based convolutional neural networks. European Journal of Clinical
1068   Microbiology & Infectious Diseases, 1-11.
1069 Singh, D., Kumar, V., Yadav, V., & Kaur, M. (2020). Deep Convolutional Neural Networks based Classification
1070   model for COVID-19 Infected Patients using Chest X-ray Images. International Journal of Pattern Recognition
1071   and Artificial Intelligence., doi: 10.1142/S0218001421510046
1072 Ting, D. S. W., Carin, L., Dzau, V., & Wong, T. Y. (2020). "Digital technology and COVID-19." *Nature medicine*
1073   26.4 (2020): 459-461.
1074 Tiwari, S., Kumar, S., & Guleria, K. (2020). Outbreak Trends of Coronavirus Disease–2019 in India: A Prediction.
1075   *Disaster medicine and public health preparedness*, 1-6.
1076 Toğaçar, M., Ergen, B., & Cömert, Z. (2020). COVID-19 detection using deep learning models to exploit Social
1077   Mimic Optimization and structured chest X-ray images using fuzzy color and stacking approaches. *Computers*
1078   *in Biology and Medicine*, 103805.
1079 Tuli, S., Tuli, S., Tuli, R., & Gill, S. S. (2020). Predicting the Growth and Trend of COVID-19 Pandemic using
1080   Machine Learning and Cloud Computing. *Internet of Things*, 100222.
1081 Tummers, J., Catal, C., Tobi, H., Tekinerdogan, B., & Leusink, G. (2020). Coronaviruses and people with intellectual
1082   disability: an exploratory data analysis. *Journal of Intellectual Disability Research*.
1083 Ucar, F., & Korkmaz, D. (2020). COVIDiagnosis-Net: Deep Bayes-SqueezeNet based Diagnostic of the Coronavirus
1084   Disease 2019 (COVID-19) from X-Ray Images. *Medical Hypotheses*, 109761.
1085 Vaid, S., Cakan, C., & Bhandari, M. (2020). Using Machine Learning to Estimate Unobserved COVID-19 Infections
1086   in North America. *JBJS*.
1087 Vaishya, R., Javaid, M., Khan, I. H., & Haleem, A. (2020). Artificial Intelligence (AI) applications for COVID-19
1088   pandemic. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*.
1089 Wang, W., Xu, Y., Gao, R., Lu, R., Han, K., Wu, G., & Tan, W. (2020). Detection of SARS-CoV-2 in different types
1090   of clinical specimens. *Jama*.
1091 Wang, Y., Dong, C., Hu, Y., Li, C., Ren, Q., Zhang, X., ... & Zhou, M. (2020b). Temporal changes of CT findings in
1092   90 patients with COVID-19 pneumonia: a longitudinal study. *Radiology*, 200843.

1093 Weidt, F., & Silva, R. (2016). Systematic literature review in computer science-a practical guide. *Relatórios Técnicos*
1094   *Do DCC/UFJF*, *1*.
1095 Wen, Z., Chi, Y., Zhang, L., Liu, H., Du, K., Li, Z., ... & Wang, D. (2020). Coronavirus disease 2019: initial detection
1096   on chest CT in a retrospective multicenter study of 103 Chinese subjects. *Radiology: Cardiothoracic*
1097   *Imaging*, *2*(2), e200092.
1098 Wong, H. Y. F., Lam, H. Y. S., Fong, A. H. T., Leung, S. T., Chin, T. W. Y., Lo, C. S. Y., ... & Lee, E. Y. P. (2020).
1099   Frequency and distribution of chest radiographic findings in COVID-19 positive patients. *Radiology*, 201160.
1100 World Health Organization. (2020). Novel Coronavirus (2019-nCoV): situation report, 3.
1101 Wu, X., Hui, H., Niu, M., Li, L., Wang, L., He, B., ... & Zha, Y. (2020). Deep learning-based multi-view fusion model
1102   for screening 2019 novel coronavirus pneumonia: a multicentre study. *European Journal of Radiology*, 109041.
1103 Wynants, L., Van Calster, B., Bonten, M. M., Collins, G. S., Debray, T. P., De Vos, M., ... & Schuit, E. (2020).
1104   Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal.
1105   *bmj*, *369*.
1106 Xie, X., Zhong, Z., Zhao, W., Zheng, C., Wang, F., & Liu, J. (2020). Chest CT for typical 2019-nCoV pneumonia:
1107   relationship to negative RT-PCR testing. *Radiology*, 200343.
1108 Xu, X., Yu, C., Qu, J., Zhang, L., Jiang, S., Huang, D., ... & Jiang, R. (2020). Imaging and clinical features of patients
1109   with 2019 novel coronavirus SARS-CoV-2. *European journal of nuclear medicine and molecular imaging*, 1-
1110   6.
1111 Yang, S., Jiang, L., Cao, Z., Wang, L., Cao, J., Feng, R., ... & Shan, F. (2020). Deep learning for detecting corona
1112   virus disease 2019 (COVID-19) on high-resolution computed tomography: a pilot study. *Annals of*
1113   *Translational Medicine*, *8*(7).
1114 Yang, Z., Zeng, Z., Wang, K., Wong, S. S., Liang, W., Zanin, M., ... & Liang, J. (2020). Modified SEIR and AI
1115   prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of Thoracic*
1116   *Disease*, *12*(3), 165.
1117 Yuen, K. S., Ye, Z. W., Fung, S. Y., Chan, C. P., & Jin, D. Y. (2020). SARS-CoV-2 and COVID-19: The most
1118   important research questions. *Cell & bioscience*, *10*(1), 1-5.
1119 Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019). Deep learning and its applications to machine
1120   health monitoring. *Mechanical Systems and Signal Processing*, *115*(January), 213–237.
1121 Zhao, Z., Chen, W., Wu, X., Chen, P. C. Y., & Liu, J. (2017). *LSTM network : a deep learning approach for short-*
1122   *term traffic forecast*. 68–75.
1123 Zu, Z. Y., Jiang, M. D., Xu, P. P., Chen, W., Ni, Q. Q., Lu, G. M., & Zhang, L. J. (2020). Coronavirus disease 2019
1124   (COVID-19): a perspective from China. *Radiology*, 200490.
1125

Figure 1

Graphical representation of the survey structure
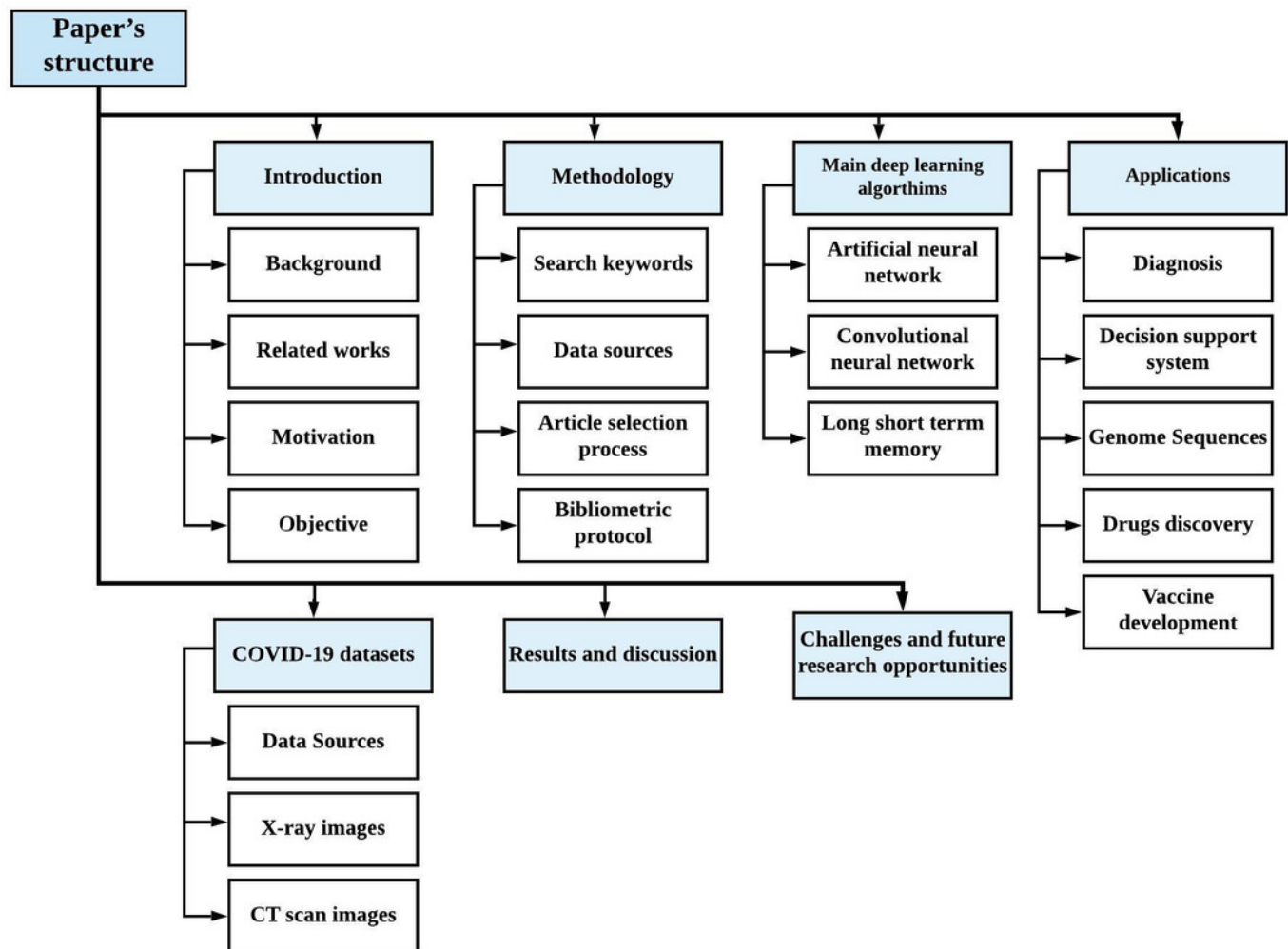
Figure 2

Article selection process
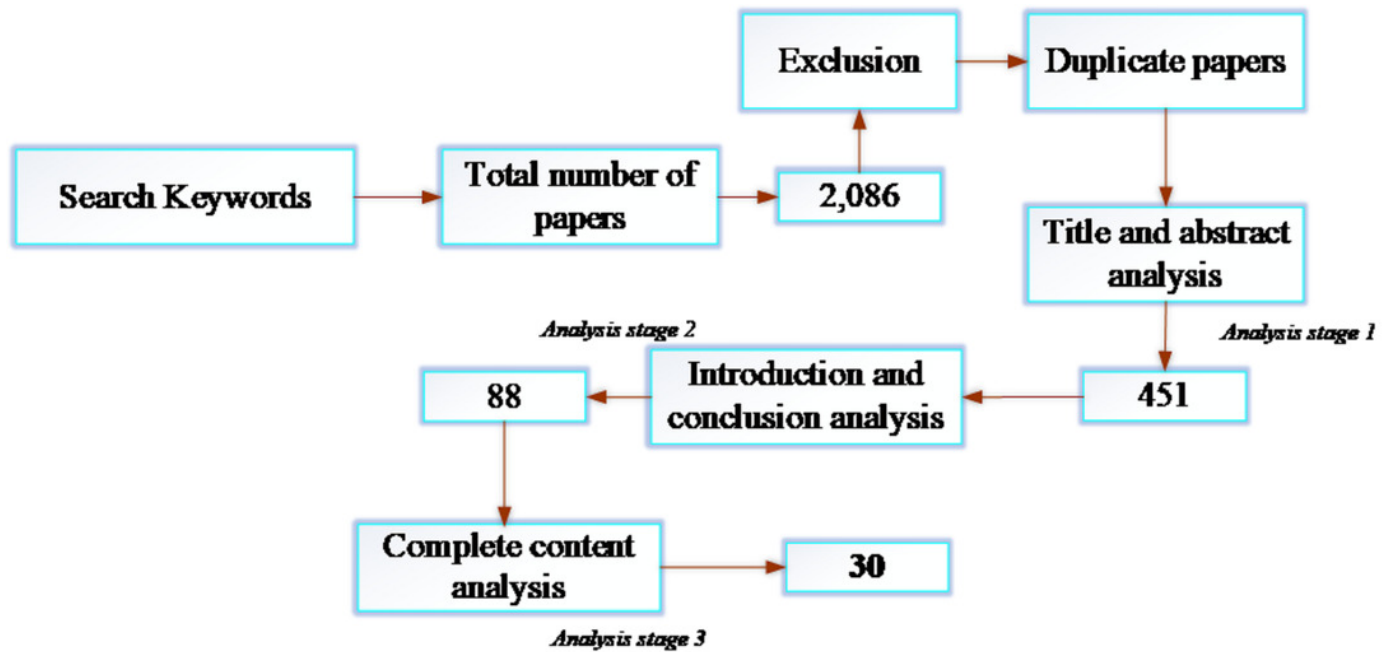
Figure 3

Taxonomy of the machine learning algorithms adopted in fighting COVID-19
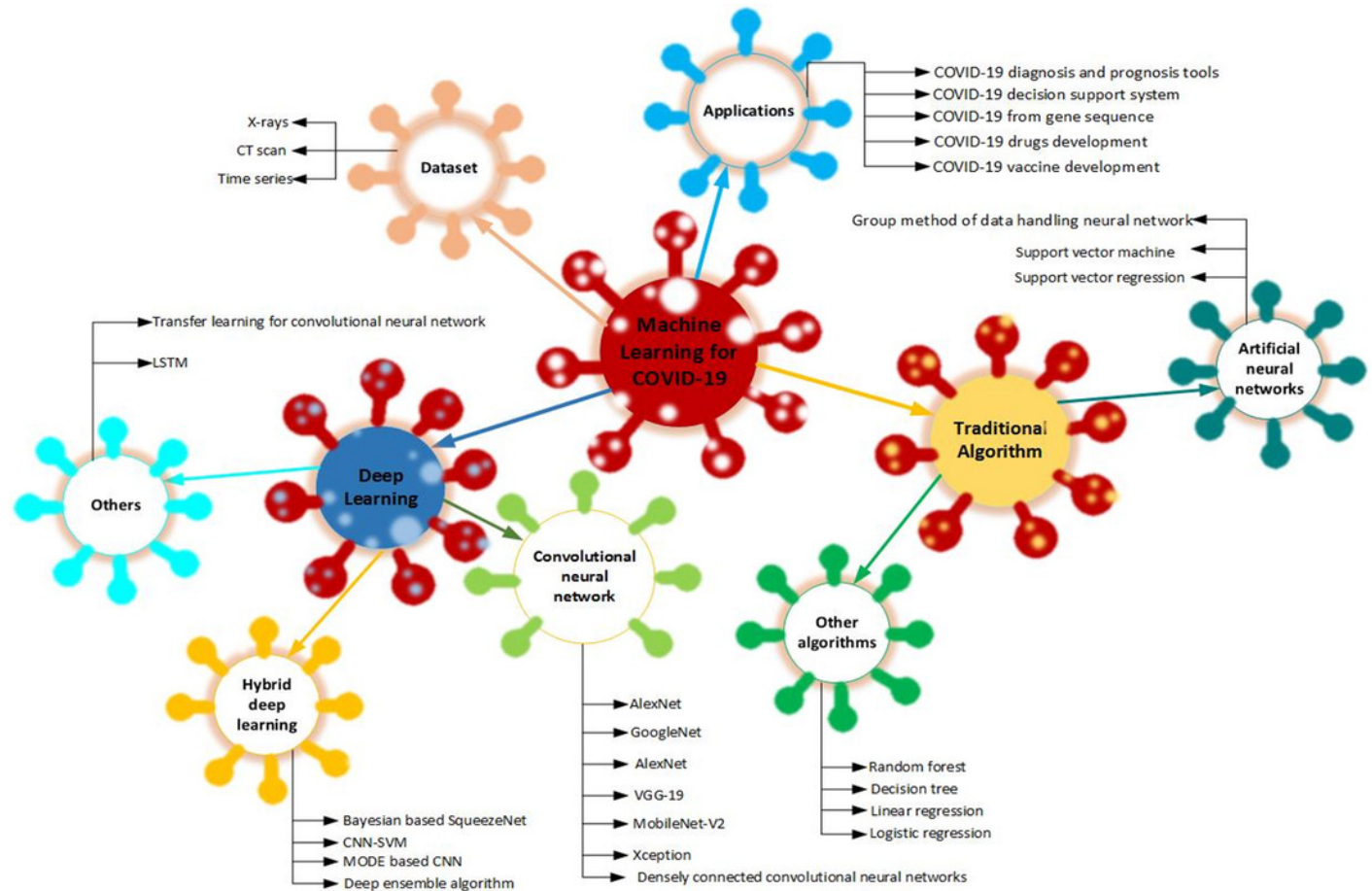
Figure 4

Visual representation of COVID-19 data extracted from different projects
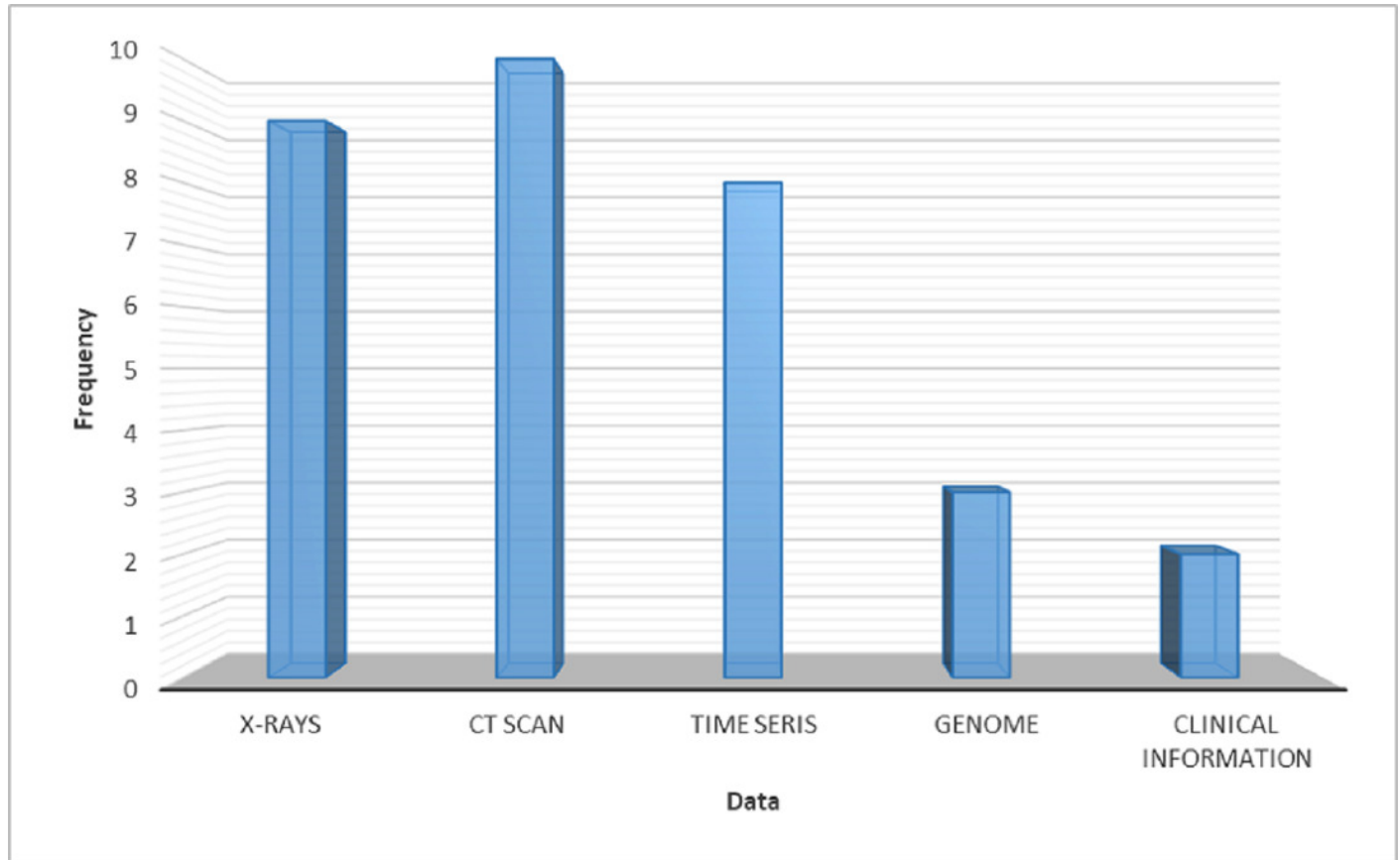
Figure 5

Machine learning algorithms adopted in fighting COVID-19
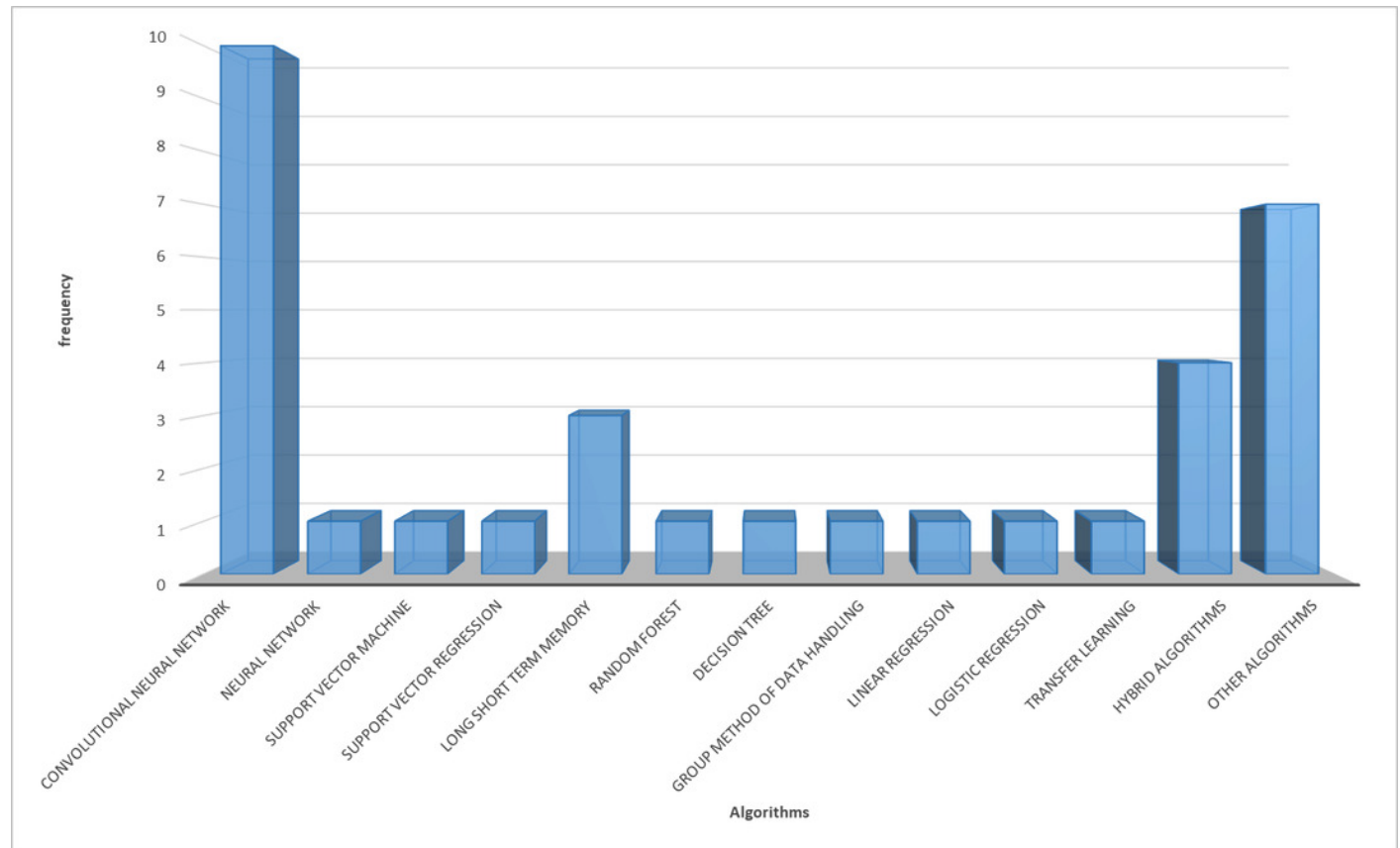
Figure 6

Different aspect of machine learning applications in fighting COVID-19 pandemic
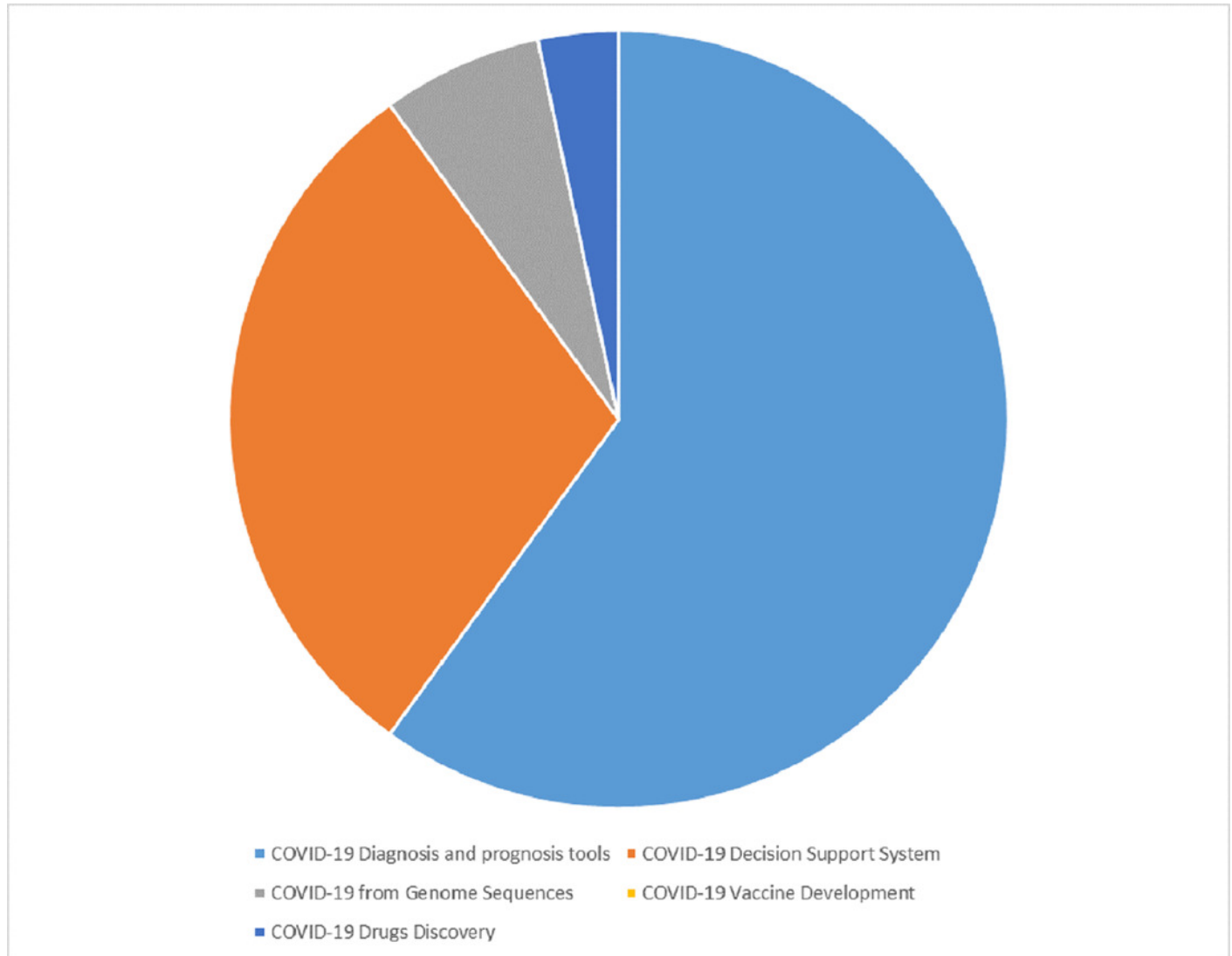
Figure 7

Trend of publications on machine learning applications in fighting COVID-19
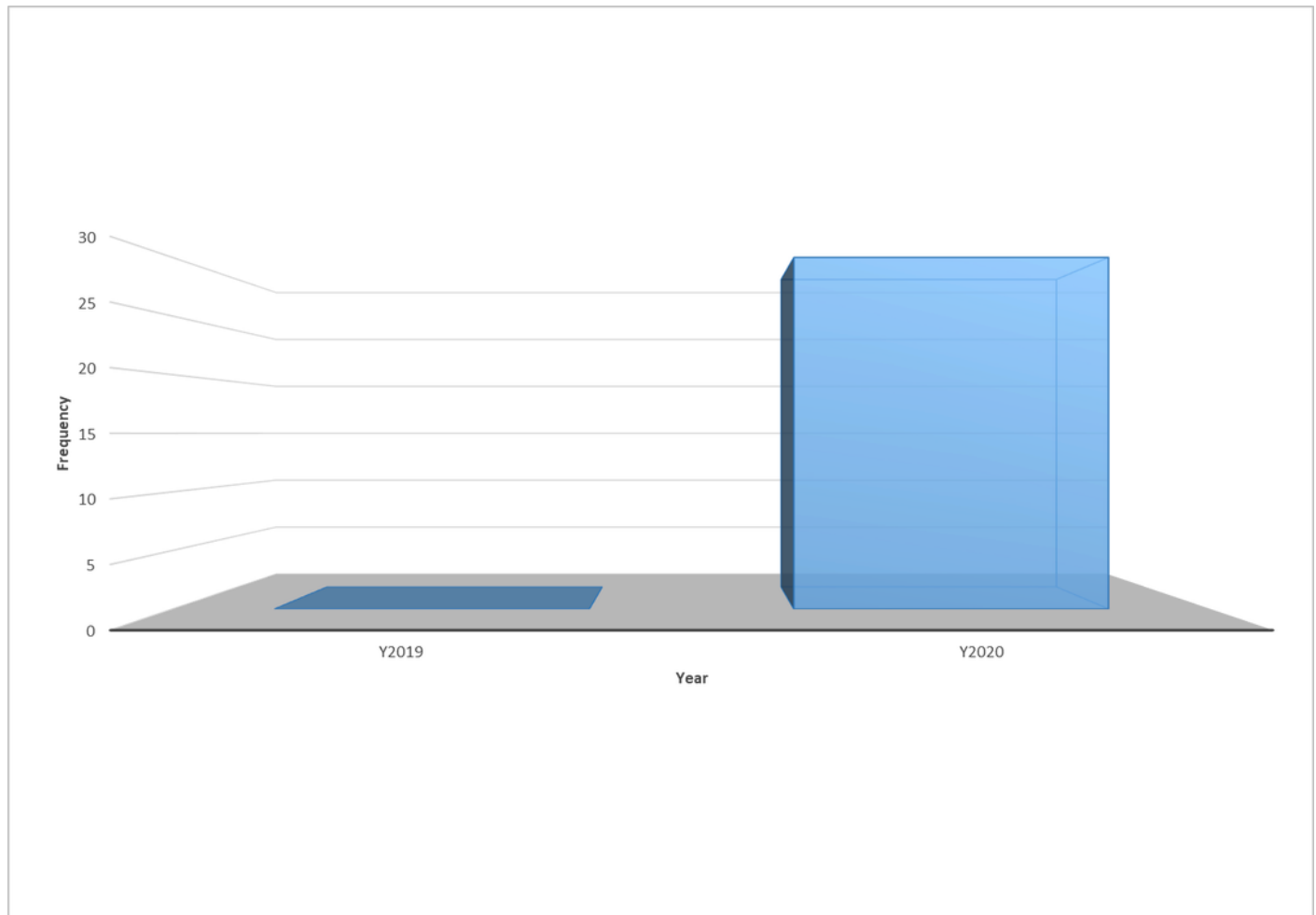
Figure 8

Bibliographic coupling among the authors
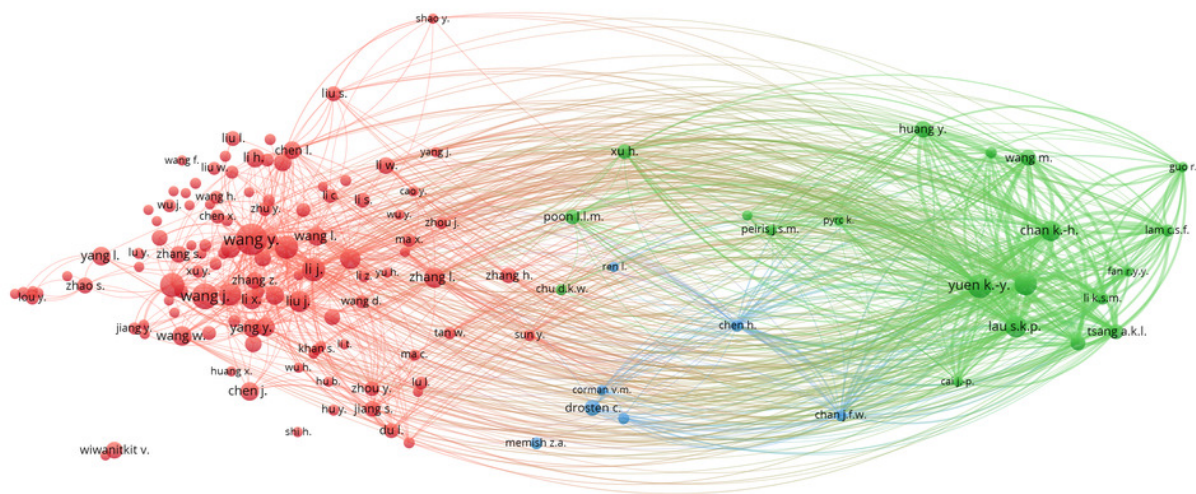
Figure 9

Bibliographic coupling among the countries
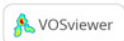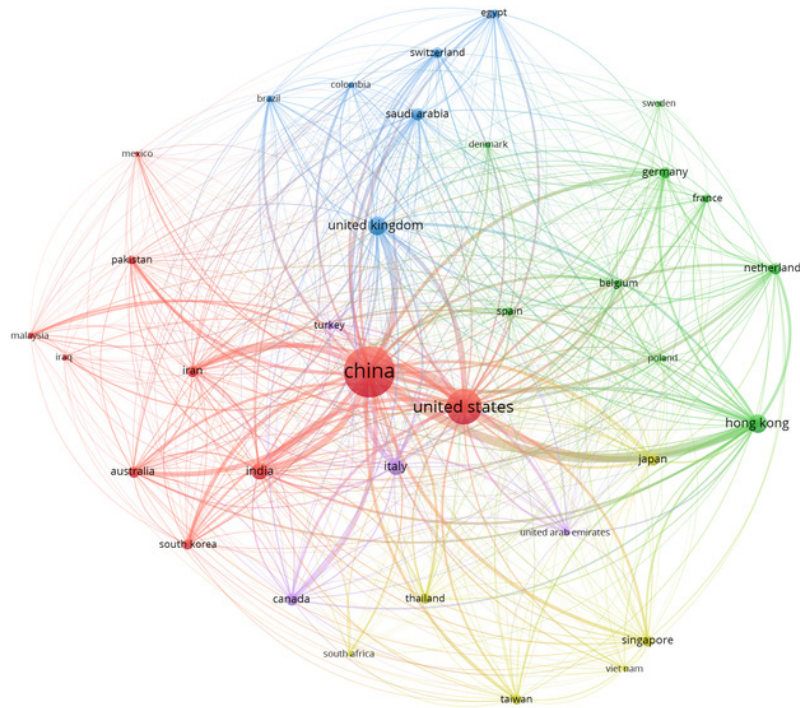
Figure 10

Bibliographic coupling among institutions
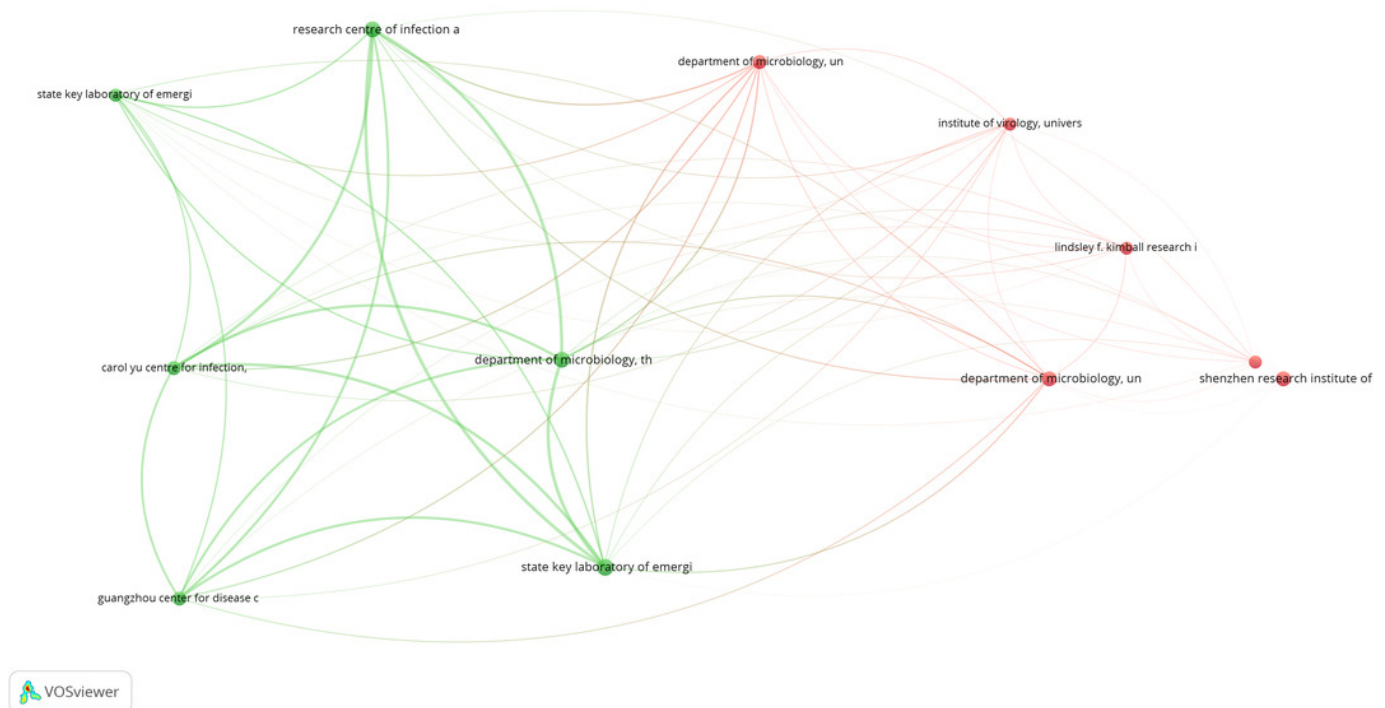
Figure 11

Bibliographic coupling among the Journals
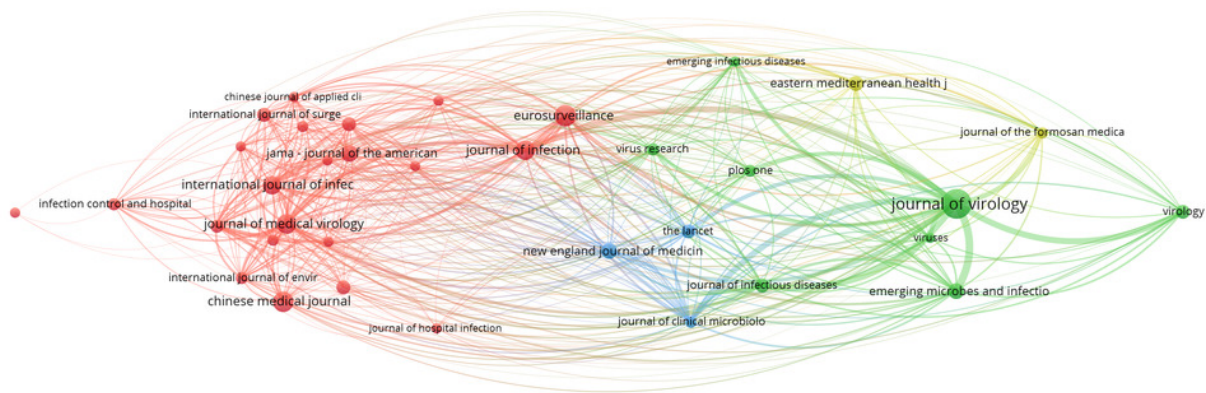
Figure 12

Co-authorship and Authors' analysis
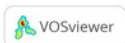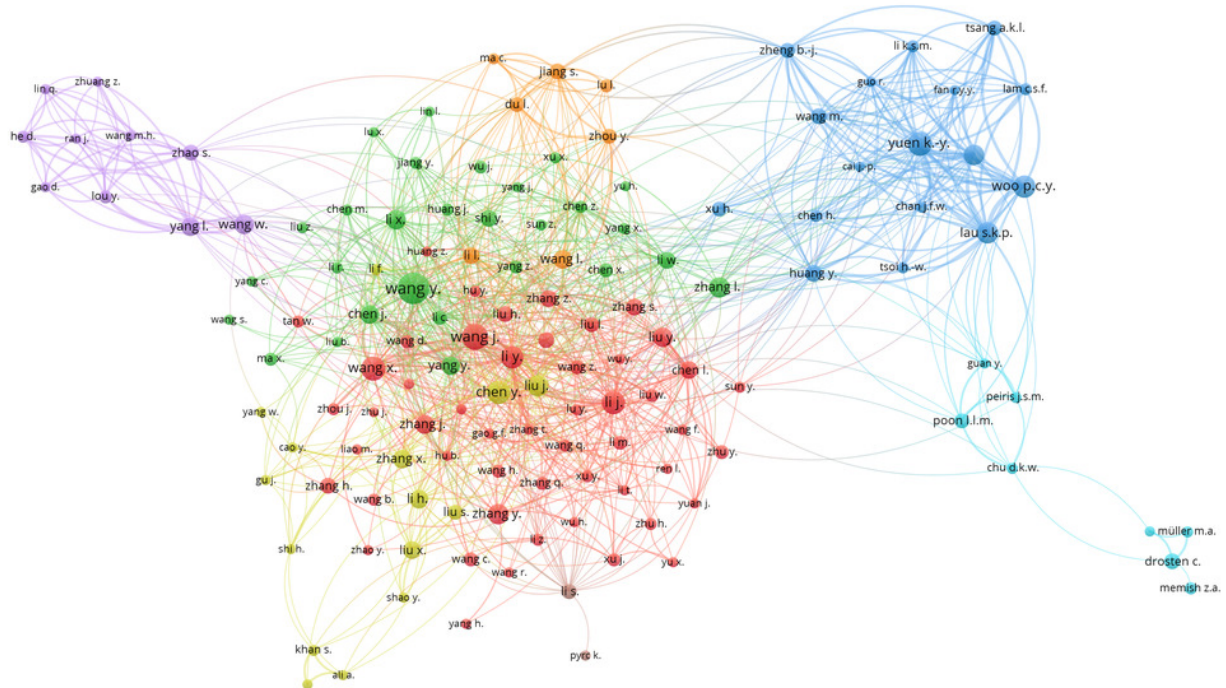
Figure 13

Author citation by institution
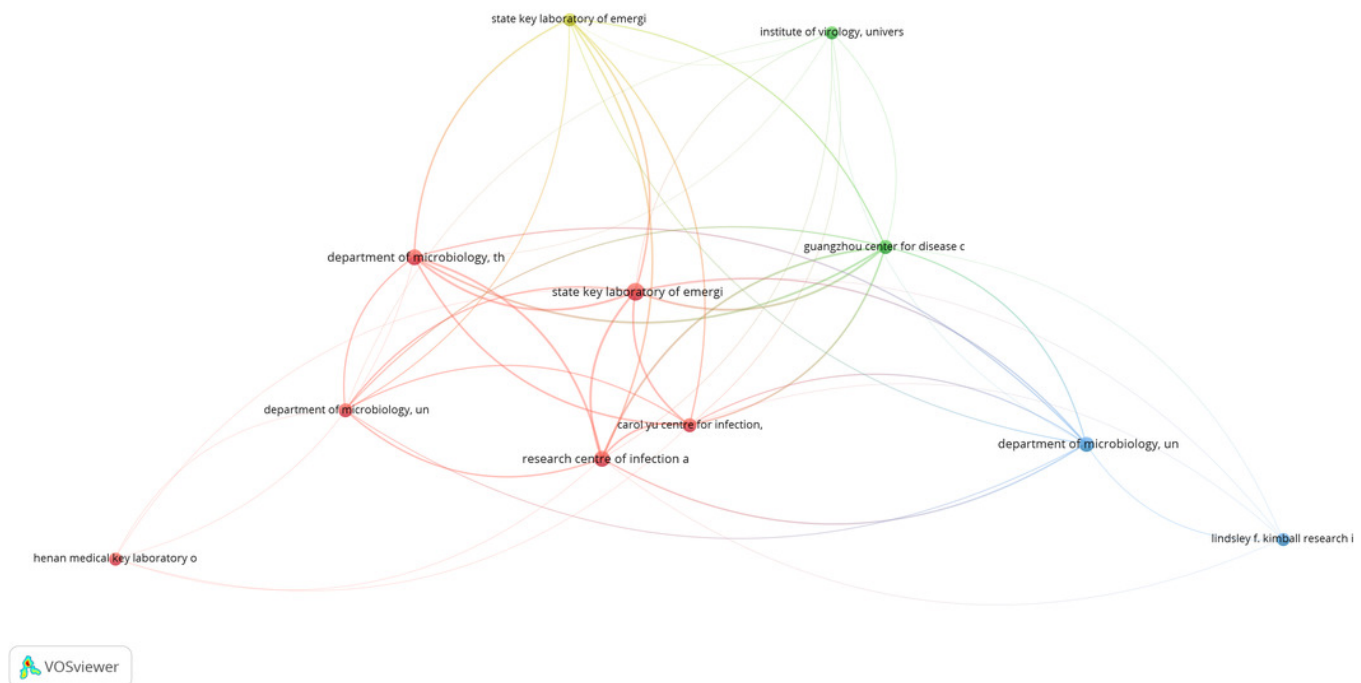
Figure 14

Author citation by journal source
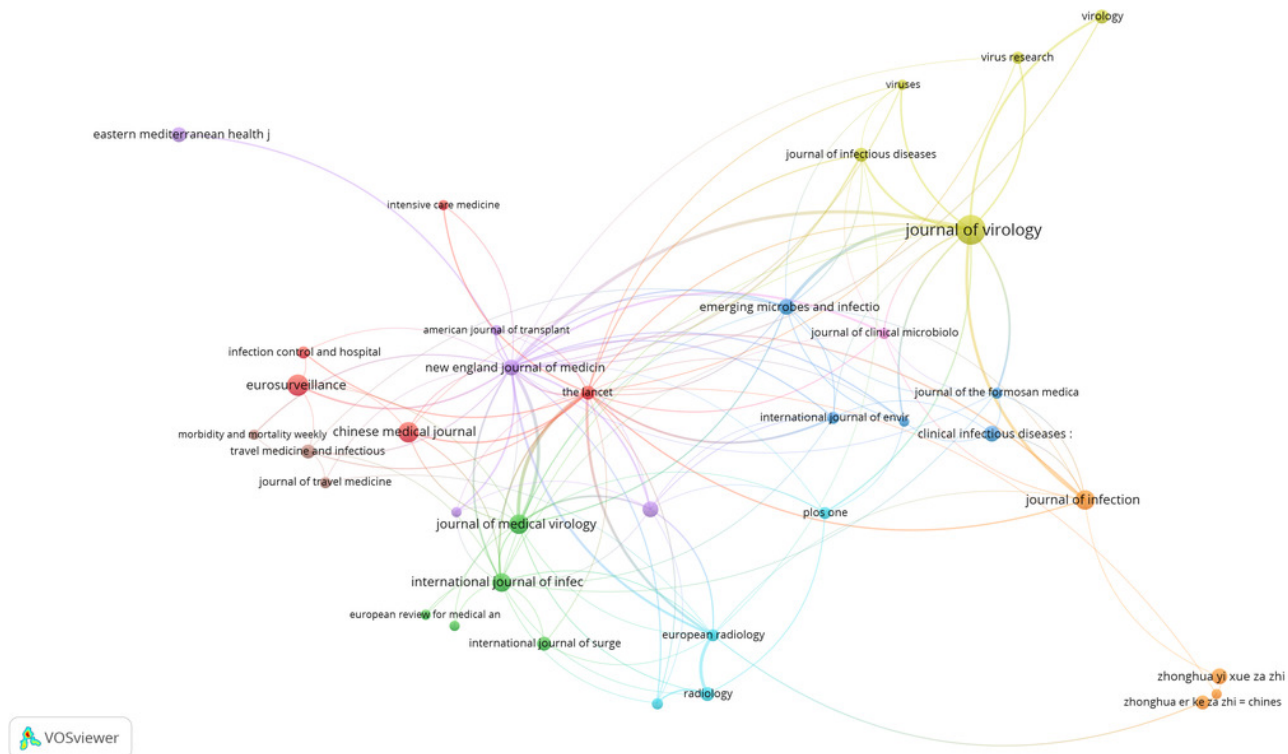
Figure 15

Visual representation of the challenges

Academic databases

Academic database	Link
DBLP	http://dblp.uni-trier.de/
ACM digital library	http://dl.acm.org/
IEEEXplore	https://ieeexplore.ieee.org/
Sciencedirect	http://www.sciencedirect.com/
Springerlink	https://link.springer.com/
PubMed	https://pubmed.ncbi.nlm.nih.gov/
Scopus	https://www.scopus.com/
Web of science	https://apps.webofknowledge.com/

Article inclusion/exclusion criteria

Inclusion criteria	Exclusion criteria
The review only focuses on COVID-19.	Other viral infections and health issues were not considered relevant in the survey.
Only articles that applied machine learning techniques to fight COVID-19 were considered.	Articles using techniques other than machine learning techniques were excluded.
Articles/conference papers published by prominent and indexed journals were included	Articles/conference papers published by non-indexed journals were excluded. The article uploaded as a preprint in preprint servers such as bioRxiv, medRxiv, arXiv, etc. without peer review were excluded.
Only articles written in the English language were considered for inclusion.	Articles written in languages other than English were excluded.

The summary of the COVID-19 diagnostic tools based on machine learning algorithms

Reference	Algorithm	Performance	Contribution	Benefit
Apostolopoulos and Mpesiana (2020)	Transfer Learning with CNN	Accuracy: **98.75%** Sensitivity: **92.85%** Specificity: **98.75%**	Device approach for automatic diagnostic of COVID-19 based on X-ray	Cost-effective, fast diagnosis and reduce exposure of medical workers to COVID-19
Ardakani et al. (2020)	Variants of CNN ResNet 101	Accuracy: **99.51%** Sensitivity: **100%** Specificity: **99.02%**	Automated characterization and diagnosis of COVID-19 infection.	Cheaper and faster compared to the traditional laboratory analysis of COVID-19. Reduces medical worker's workload.
Butt et al. (2020)	CNN	AUC: **0.996** Sensitivity: **98.2%** Specificity: **92.2%**.	Outperform the traditional RT-PCR testing of COVID-19	fast and reliable in detecting COVID-19 pandemic.
Huang et al. (2020)	Deep learning algorithm	Not Applicable as a result of ANOVA analysis	The assessment of the lung opacification measured is significantly different from the clinical groups	The approach has the potential to remove the subjectivity in the initial evaluation of COVID-19 findings as well as follow up pulmonary
Li et al. (2020)	CNN	Sensitivity:**87%** Specificity:**92%** AUC: **95%**	The automated framework that differentiates COVID-19 from pneumonia	Automated the COVID-19 testing process, reduces the testing time and fatigue.
Liu et al. (2020)	P-CNN	Sensitivity: **100%** Precision: **76.9%**	Diagnose COVID-19 with limited data and present new probabilistic Grad-CAM salient map	Limited amount of data can be used for the COVID-19 diagnoses, and it is interpretable.
Ozturk et al. (2020)	CNN	Sensitivity: **85.35%** Specificity: **92.18%** Accuracy: **87.02%**	Improve efficiency and automate the COVID-19 screening process	Automate the process of COVID-19 diagnoses to reduce fatigue
Salman et al. (2020)	CNN	Sensitivity: **100%** Specificity: **100%** Accuracy: **100%**	Automated COVID-19 screening process	Reduces diagnostic time
Singh et al. (2020)	MODE based CNN	Sensitivity: **> 90%** Specificity: **> 90%**	Diagnose COVID-19 with better accuracy than the competitive models	The proposal is beneficial to COVID-19 real-time classification
Togaçar et al. (2020)	CNN-SVM	Overall accuracy: **99.27%**	Contributed to the efficient detection of COVID-19	Automated detection of COVID-19 patient
Ucar and Korkmaz (2020)	Bayesian-based SqueezeNet	Accuracy: **100%** Specificity: **99.67%**	Presents alternative rapid COVID-19 diagnostic tool based on deep Bayes-SqueezeNet	It will be of benefit to healthcare professionals in diagnosing COVID-19 efficiently.
Wu et al. (2020)	ResNet50	Accuracy: **0.819** Sensitivity: **0.760** Specificity: **0.811**	Provide COVID-19 diagnosis from multiple view	Reduce the workload of a radiologist by offering fast and accurate COVID-19

			diagnosis	
Loey et al. (2020)	Googlenet, Alexnet and RestNet18	Googlenet, Alexnet and Googlenet scored **80.6%, 85.2%** and **100%** in the 4, 3 and 2 classes classification problems	Generate sufficient COVID-19 data to improve COVID-19 diagnosis	Early detection of COVID-19 and reduce the workload of radiologist
Hurt et al. (2020)	Deep learning	Not provided	Improve the detection COVID-19 from x-ray	Early detection of COVID-19
Jiang et al. (2020)	Machine learning algorithm	Accuracy: **70% - 80%**	Detect COVID-19 severity in a patient at the initial presentation	Help in optimal utilization of scarce resources to cope with COVID-19
Liu et al. (2020)	Logistic regression	ROC: **0.93** Confidence interval: **95%**	Applied CT quantification of pneumonia to predict progression to COVID-19 severity	provide a prognostic indicator for COVID-19 clinical management
Mei et al. (2020)	Deep ensemble algorithm	ROC: 0.92 Accuracy: **68%** Sensitivity: **84.3%** Specificity: **82.8%**	Predict COVID-19 with both image and none image clinical information	The ensemble diagnostic tool can detect COVID-19 patients rapidly
Yang et al. (2020b)	Densely connected convolutional networks	Accuracy: **92%** Sensitivity: **97%** Specificity: **87%**	Detect COVID-19 from CT scan via densely connected convolutional networks	Reduce radiologist workload

Summary of the adoption of machine learning approaches in building COVID-19 decision support systems

Reference	Algorithm	Performance	Contribution	Benefit
Ayyoubzadeh et al. (2020)	LSTM & Linear regression	LSTM: RMSE **27.187** Linear Regression: RMSE **7.562**	Predict COVID-19 positive cases in Iran	The algorithm can predict the trend of the COVID-19 pandemic in Iran, which can help policymakers to plan for the allocation of medical resources.
Chimmula & Zhang (2020)	LSTM	RMSE: **34.83** Accuracy: **92.6%**	Forecast COVID-19 transmission in Canada	Help decision-makers in monitoring and curtailing future transmission of COVID-19 in Canada
Liu et al. (2020b)	ANN	Not Applicable	Estimated the trend of COVID-19 in China	Help policymakers and health officials attend to the need of other diseases during the COVID-19 pandemic
Ribeiro et al. (2020)	Support vector regression	MAE: **79.17**	Provide future COVID-19 confirmed cases in brazil	monitoring COVID-19 cases in Brazil and help decision-makers in taken critical decision about COVID-19
Tiwari et al. (2020)	Machine learning algorithm (not specified)	MAE & RSME – graphical	The predicted peak period of COVID-19 in India	Help India policymakers decide on COVID-19 to mitigate its spread
Tuli et al. (2020)	Machine learning algorithm (not specified)	MSE: **9.32E+06**	Provide real life COVID-19 predictions	Government and citizens can use the results for proactive measures to fight COVID-19
Vaid et al. (2020)	Machine learning algorithm (not specified)	Not reported	Predict Potential COVID-19 infections	Policymakers in North America can use the projection to curtail the effect of COVID-19 pandemic
Yang et al. (2020)	LSTM	Confidence Interval: **95%**	Predict COVID-19 trend in China	Authorities in China to decide to control the COVID-19 pandemic
Pirouz et al. (2020)	Group method of data handling neural network	Accuracy: **85.7%**	Predict COVID-19 pandemic based on weather condition	Help in managing COVID-19 pandemic

The summary of detecting COVID-19 from genome sequence via an algorithm

Reference	Algorithm	Performance	Contribution	Benefit
Qiang et al. (2020)	Random Forest	Accuracy: **98.18%**	Able to detect none human COVID-19 origin from spike protein	used in COVID-19 genome mutation surveillance
Randhawa et al. (2020)	Decision Tree	Mathew Correlation Coefficient: **0.9638** Accuracy: **100%**	Successfully used intrinsic viral genomic signatures to classify COVID-19 with 100% accuracy	The DT-DSP is a reliable real-time alternative for the classification of taxonomic

Summary of COVID-19 data sources and accessibility

Reference	Data	Owners	Source/ accessibility	Remarks
Apostolopoulos and Mpesiana (2020); Ozturk et al. (2020); Salman et al. (2020); Togaçar et al. (2020); Ucar and Korkmaz (2020); Loey et al. (2020)	X-Ray	Cohen	https://github.com/ieee8023/covid-chestxray-dataset	A project with a collection of x-rays images
Apostolopoulos and Mpesiana (2020)	X-Ray	Kaggle	https://www.kaggle.com/andrewmvd/convid19-X-rays	Dataset_1: 224 images with Covid-19, 700 images with common bacterial pneumonia, and 504 normal images condition. Dataset_2: 224 cases of Covid-19, 504 healthy instances, and 714 has both bacterial and viral pneumonia.
Ardakani et al. (2020)	CT scan	Ardakani et al. (2020)	16-MDCT scanner (Alexion, Toshiba Medical System, Japan)	The data contained 1020 CT slices from 108 patients confirmed COVID-19 and 86 patients without COVID-19.
Ayyoubzadeh et al. (2020); Vaid et al. (2020)	Time series	Worldometer	https://www.worldometers.info/coronavirus/	The daily new COVID-19 cases from 02/15, 2020, to 03/18/ 2020 in Iran. Dataset features: previous day's search trends, previous day cases, and Output: new cases of the current day.
Barbosa & Fernandes (2020)	Genome	Barbosa & Fernandes (2020)	https://data.mendeley.com/datasets/nvk5bf3m2f/1	The data is chaos game representation of SARS-CoV-2 containing both the raw and processed data with 100 instances of SARS-CoV-2 genome
Butt et al. (2020)	CT scan	Butt et al. (2020)	Butt et al. (2020)	The data comprised of 618 CT scan samples including 219 from 110 COVID-19 patients.
Chimmula & Zhang (2020)	Time series	Johns Hopkins University and Canadian Health authority	Chimmula & Zhang (2020)	The data is available in time series: date, month and year up to 31/04/20
Huang et al. (2020)	CT scan	Huang et al. (2020)	Huang et al. (2020)	126 COVID-19 patients that underwent a CT chest scan from 1/1/2020 to 3/2/ 2020
Li et al. (2020)	CT scan	Li et al. (2020)	Li et al. (2020)	Data from 3,322 patients

				comprising of 4356 chest CT scans. Data collection period from 16/08/ 2016 to 17/02/2020
Liu et al. (2020)	Time series	1.Tencent news 2.Baidu	Real COVID-19 reporting: https://news.qq.com/zt2020/page/feiyan.htm?ADTAG=area Baidu: http://qianxi.baidu.com	1. COVID-19 time-series data including locations confirmed cases, deaths, recovery cases, and new diagnosed cases
				2. Baidu is an open-source for big data project that provides visualization of population migration
Oh et al. (2020)	X-ray	Society of Radiological Technology	Oh et al. (2020)	Segmentation network dataset with a total of 247 posteroanterior chest radiographs
Qiang et al. (2020)	Protein sequence	China national genomic data center	https://bigd.big.ac.cn/ncov	The database has an extensive protein sequence
Randhawa et al. (2020)	Genome	National Center for Biotechnology Information	https://sourceforge.net/projects/mldsp-gui/files/COVID19Dataset	The database has genome dataset with thousands of *bp*
Ribeiro et al. (2020)	Time series	Ribeiro et al. (2020)	https://brasil.io/api/dataset/covid19/caso/data/?place_type=state	Daily COVID-19 reports on confirmed cases from different states in Brazil
Singh et al. (2020)	CT scan	Singh et al. (2020)	Singh et al. (2020)	Chest CT scan images from ICU
Tiwari et al. (2020)	Time series	Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU).	https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset	Confirmed COVID-19 cases, recovered cases, and death cases from China
Tuli et al. (2020)	Time series	Hannah Ritchie	https://ourworldindata.org/coronavirus-source-data	The dataset is the Our World in Data provided by Hannah Ritchie
Wu et al. (2020)	CT scan	Wu et al. (2020)	Wu et al. (2020)	The CT scan images of 495 patients were collected from 3 hospitals in China
Yang et al. (2020)	Time series	1. National health commission of China 2. Baidu 3. SOHU	1. http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml 2. https://qianxi.baidu.com/ 3. http://news.sohu.com/57/26/subject206252657.shtml	SARS and COVID-19 datasets
Hurt et al. (2020)	x-ray	Hurt et al. (2020)	Hurt et al. (2020)	Chest x-ray images from China and America

Jiang et al. (2020)	CT scan & clinical characteristics	Jiang et al. (2020)	Jiang et al. (2020)	CT scan images and clinical characteristics
Liu et al. (2020)	CT scan	Liu et al. (2020)	Liu et al. (2020)	Chest CT scan that was performed using a 64-slice CT scanner without contrast agents.
Mei et al. (2020)	CT scan and clinical information	Mei et al. (2020)	Mei et al. (2020)	Both clinical information and chest CT scan images were collected for the study
Pirouz et al. (2020)	Time series	Pirouz et al. (2020)	Pirouz et al. (2020)	Environmental and urban data
Yang et al. (2020b)	CT scan	Yang et al. (2020b)	Yang et al. (2020b)	CT scan images from a hospital in china