



# A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of COVID-19

Yousseoufa Mohamadou<sup>1,2</sup> · Aminou Halidou<sup>3</sup> · Pascal Tiam Kapen<sup>2,4,5</sup>

© Springer Science+Business Media, LLC, part of Springer Nature 2020

## Abstract

In the past few months, several works were published in regards to the dynamics and early detection of COVID-19 via mathematical modeling and Artificial intelligence (AI). The aim of this work is to provide the research community with comprehensive overview of the methods used in these studies as well as a compendium of available open source datasets in regards to COVID-19. In all, 61 journal articles, reports, fact sheets, and websites dealing with COVID-19 were studied and reviewed. It was found that most mathematical modeling done were based on the Susceptible-Exposed-Infected-Removed (SEIR) and Susceptible-infected-recovered (SIR) models while most of the AI implementations were Convolutional Neural Network (CNN) on X-ray and CT images. In terms of available datasets, they include aggregated case reports, medical images, management strategies, healthcare workforce, demography, and mobility during the outbreak. Both Mathematical modeling and AI have both shown to be reliable tools in the fight against this pandemic. Several datasets concerning the COVID-19 have also been collected and shared open source. However, much work is needed to be done in the diversification of the datasets. Other AI and modeling applications in healthcare should be explored in regards to this COVID-19.

**Keywords** COVID-19 · Corona virus · Mathematical modeling · Artificial intelligence · Open source dataset

## 1 Introduction

The World Health Organization declared that new coronavirus disease 2019 (COVID-19) was a Public Health

Emergency of International Concern on January 30th 2020 [1, 2]. By then there were a total number of 7818 confirmed cases of COVID-19 globally with more than 1370 severe cases and 170 deaths. The bulk of which was found in China [3]. Over the course of a few weeks the disease has propagated across the boundaries of China infecting nearly every country. At the time of writing this paper (May 01, 2020) there is a total of 2,397,216 confirmed cases globally with 162,956 deaths [4]. Symptoms of the disease include dry cough, sore throat, and fever. Although the majority of the cases are mild, some cases could lead to Acute Respiratory Distress Syndrome (ARDS), severe pneumonia, pulmonary oedema, and organ failure [5]. After the emergency declaration of WHO, several works have been done in the terms of modeling and prediction to try and provide ways to either understand the disease propagation, evaluate preventive measure put in place by authorities, provide early and accurate detection of the disease just to name a few. Mathematical modeling has been used for several years in epidemiological studies [6]. Mathematical modeling of disease transmission and propagation helps in the prediction of the course of epidemics, the design of mass vaccination

---

This article belongs to the Topical Collection: *Artificial Intelligence Applications for COVID-19, Detection, Control, Prediction, and Diagnosis*

---

✉ Yousseoufa Mohamadou  
usufcom@hotmail.com

<sup>1</sup> University Institute of Technology, University of Ngaoundere, P.O Box 454, Ngaoundere, Cameroon

<sup>2</sup> BEEMO Lab, ISST, Université des Montagnes, P.O. Box 208 Bangangté, Cameroon

<sup>3</sup> Department of Computer Science, University of Yaounde I, 812, Yaounde, Cameroon

<sup>4</sup> URISIE, University Institute of Technology Fotso Victor, University of Dschang, P.O Box 134, Bandjoun, Cameroon

<sup>5</sup> UR2MSP, Department of Physics, University of Dschang, P.O Box 67, Dschang, Cameroon

programs and also it can provide guidance on what type of data are relevant in the study of the epidemics [7]. Some of the studies carried out in regards to the current COVID-19 include modeling of the dynamic of COVID-19, exploring the effect of prevention method like travel restriction of COVID-19 and studying the effect of climate on the COVID-19 propagation [8]. On the other hand, artificial intelligence (AI) is a tool used for prediction. AI is the study and development of algorithms (machines) that mimic human intelligence. AI has been successfully used in a several fields such as computer vision, online advertising, spam filtering, robotics, fraud detection and so on [9, 10]. In healthcare, AI has also gained attention in terms of disease detection, treatment selection, patient monitoring, drug discovery, gene function annotation, automated experiments, automated data collection etc. [11, 12]. As to what concerns the COVID-19, AI has been used in medical image acquisition, image segmentation and diagnosis [13]. In this paper, a review of the mathematical modeling and artificial intelligence used in the study, estimation and prediction of COVID-19 is presented. The paper is divided into three parts, the first presents the mathematical models used in the study of the pandemic, the second presents the various AI applications in disease diagnosis and estimation and in the third part a list of available datasets for COVID-19 is presented.

## 2 Material and method

The review is divided into three parts each dealing with a specific aspect like Mathematical modeling, AI applications and available datasets. For each of the three parts, the items reviewed were grouped into topics and then a summary of each group is done. In all a total number of 61 journal articles, reports, fact sheets and websites were reviewed. The items reviewed were all published between December 2019 to April 2020. Table 1 shows the structure of the review including the number of items reviewed and the main focus of the reviewed items.

## 3 Mathematical modeling and COVID-19

Various research works were developed in literature for the modeling of dynamics and spread of COVID-19. Most of these were particularly based on the Susceptible-Exposed-Infected-Removed (SEIR) model and the Susceptible-infected-recovered (SIR) model. These models were largely used in the past for the study of epidemic spreading with various forms of networks of transmission [14–22]. Table 2 gives the summary of the various models used in COVID-19 studies. The following gives a review of these models.

### 3.1 Susceptible-exposed-infected-removed (SEIR)

Choujun et al. [23] used daily intercity migration data together with a SEIR model to generate a new model that describes the dynamics of COVID-19 in China. They collected the daily intercity migration data from 367 cities using a mobile application that tracks human migration. They concluded that the number of infections in most cities in China would be highest between the middle of February to early March 2020. Anca and Kieran adapted a traditional SEIR model to study the specific dynamic compartments and epidemic parameters of COVID-19 [24]. They analyzed the current management strategy of the pandemic, including social distancing, travel bans, and service interruptions and closures for the generation of predictions, and assessment of the efficiency of these control measures. In [25], the combination of SEIR and regression models was used with John Hopkins University dataset on COVID-19 for the prediction of the change in the spreading of COVID-19. The study presented in [26] used an age-structured susceptible-exposed-infected-removed (SEIR) model for physical distancing measurement and evaluation. The authors showed that physical distancing measures were most effective if the gradual return to work started in April. The study of the transmission of the COVID-19 and its association with temperature and humidity using the SEIR model was initiated by Xiao-Jing et al. [27]. The outcomes of the study presented that raising the temperature and humidity values contributed to the control of transmission of the disease. In [28], the SEIR model was adapted to investigate the potential community-wide impact of public use of face masks on the transmission dynamics and control of the COVID-19 pandemic. It was suggested that face masks should be used nation-wide and implemented immediately (Table 3).

### 3.2 Susceptible-lovered (SIR)

A time-dependent susceptible-infected-recovered (SIR) model to track the transmission rate and the recovering rate at a particular time was proposed in [29]. They obtained a prediction error of 3% or less for confirmed cases and predicted that the day the recovering rate over took the transmission rate was on February 17, 2020 in the Hubei province of China. Wang et al. [30] modified the SIR model by adding different types of time-varying quarantine strategies such as government imposed mass isolation policies and micro-inspection measures at the community level to establish a method of calibrating cases of under-reported infections. The SIR model was also used to fit the cumulative data of COVID-19 to an empirical form in China [31]. It was reported that for given parameter values, the SIR model on the Euclidean network obtained

**Table 1** The breakdown of the review showing number of items covered per part

Part	Items	Description
Mathematical Modeling	19	Modeling of CoVID-19 epidemic dynamics and propagation, Climate and environment effect on CoViD-19 spreading, Modeling of the effect of management strategies on COVID-19 spreading
Artificial Intelligence	18	Image based (X-ray, CT) AI CoViD-19 detection and classification, Text based AI CoViD-19 detection and classification, physiological data based AI CoViD-19 detection and classification
Datasets	24	Prevalance rate, medical images, Text (social media), demography

high accuracy on data form China and predict when the pandemic would be expected to be over. In [32], a simple age-sensitive SIR model, which integrated known age-interaction contact patterns for the examination of potential effects of age-heterogeneous mitigations on an epidemic in a COVID-19-like parameter regime was studied. Authors found that strict age-targeted mitigation strategies had the potential to reduce mortalities. The age-structured SIR model with social contact matrices and Bayesian imputation was studied to evaluate the progress of the pandemic in India [33]. The authors evaluated the influence of social distancing measures like workplace non-attendance, and school closure on the transmission of the novel Corona virus. It was found that a three-week lockdown would be insufficient to prevent the spread of the disease. A

simple SIR model modified to include certain variables of containment measures taken worldwide was used to study these measures [34]. By comparing various scenarios, it was shown that the infection progress strongly affected by the measures taken.

### 3.3 Other models

A Susceptible-Infectious-Quarantined-Recovered (SIQR) model for the analysis of data in Brazil was used [35]. It was found that the number of quarantined individuals grew exponentially and stabilized. The SEIQR (Susceptible-Exposed-Infectious-Quarantined-Recovered) model with time delays for latency and an asymptomatic phase was investigated [36]. It was reported that time-varying social

**Table 2** Summary of the various mathematical models used in COVID-19 studies

No.	Model	COVID-19 studies	References
1	Susceptible-Exposed-Infected-Removed (SEIR)	Dynamics, prediction, management strategies, Effect of temperature and humidity levels	[23–28]
2	Susceptible-infected-recovered (SIR)	Track transmission and recovering rates in time, data fitting, management strategies	[29–34]
3	Susceptible-Infectious-Quarantined-Recovered (SIQR)	Quarantine, management strategies	[35]
4	Susceptible-Exposed-Infectious-Quarantined-Recovered (SEIQR)	Prediction, management strategies	[36]
5	Bats-Hosts-Reservoir-People transmission network (BHRP)	Simulate transmission from the bats to human	[37]
6	Susceptible-Exposed-Symptomatic-Asymptomatic-Recovered-seafood Market (SEIARW)	Age dependent Transmissibility, prediction	[38]
7	Markov Chain Monte Carlo (MCMC)	Effects of self-protective measures	[39]
8	SPSS modeler	Effect of temperature levels	[40]
9	ODE metapopulation model	COVID-19 and economics	[41]

**Table 3** Summary of the classifications methods used in COVID-19 studies

AI	Methods	Images	Patients	Dataset	Accuracy	Sensitivity	PPV
CNN	COVID-Net [49]	13,800	13,725	COVIDx test	92.6%	87.1%	96.4%
CNN	ResNet50 ; InceptionV3 ; Inception-ResNetV2 [50]	100	50	GitHub	98% ; 97% ; 87%	–	–
CNN	COVNet [51]	4356	3,322	proprietary		87%	–
CNN	Deep Learning with X-ray [52]	448	–	proprietary	96.78%	98.66%	–
CNN	COVIDX-Net (VGG19 and DenseNet201) [53]	50	25	proprietary	90%	–	–
SVM	Barstugan [54]	150	53	proprietary	99.68%	–	–
SVM	ResNet50 and SVM [55]	158	–	GitHub, Kaggle and Open-i	95.38%	97.29%	–
SVM	SVM and Random Forests [56]	–	235	Hospital Israelita Albert Einstein in São Paulo	84.7%	6.77%	77.8%
SVM	MLT and SVM [57]	40		Montgomery County X-ray Set and covid-chestxray-dataset-master	97.48%	95.76%	99.7%
LR	Kunhua [58]	–	83	proprietary	87%	80%	82.8%
LR	SMOTE [59]	5840	88	Chest X-Ray Images (Pneumonia)1 and COVID-19 public dataset from Italy	96.6%	96.7%	98.3%
NB	Probabilistic Model [60]	51		Kaggle benchmark dataset	99.4%	–	–
LDA	NLR&RDW-SD [61]	–	45	Jingzhou Central Hospital	85.7%	90.0%	84.7%
DT/RF	RF based model [62]	–	176	proprietary	87.5%	–	93.3%
DT/RF	SMOTE [59]	5840	88	Chest X-Ray Images (Pneumonia)1 and COVID-19 public dataset from Italy	93.1%	93.2%	96.5
DT/RF	iSARF [63]	–	1658	3 University Hospitals (Tongji, Shanghai, Fudan)	87.9%	90.7%	–
k-NNA	SMOTE [59]	5840	88	Chest X-Ray Images (Pneumonia)1 and COVID-19 public dataset from Italy	94.7%	94.7%	97.4%
U-Net	Modified U-Net structure [64]	110	60	SIRM	79%	–	83%
U-Net	Attention U-Net with an adversarial critic model [65]	1047	641	JSRT, Montgomery, and Shenzhen	96%	–	–
InfNet	InfNet and the Semi-Inf-Net [66]	1600	–	CCOVID-19 CT Segmentation and COVID-19 CT/X-ray Collection	–	72.5%	–

distancing, using the SEIQR model, could reduce the number of infections by about 50%.

Recently, a novel model known as Bats-Hosts-Reservoir-People transmission network model was used to simulate the potential transmission from bats (infection source) to human [37]. Another method was developed where the age-specific Susceptible-Exposed-Symptomatic-Asymptomatic-Recovered-Seafood Market (SEIARW) model based on two suspected transmission routes was used to quantify age-specific transmission

[38]. The two routes were from market to person and from person to person. The authors concluded that COVID-19 transmissibility is higher in elderly persons as compared to young persons. In [39] the influence of interventions and self-protection measures (travel restriction, quarantine of entry, contact tracing, isolation and wearing masks) on COVID-19 transmission dynamic in mainland China excluding Hubei province was modeled using the Markov Chain Monte Carlo (MCMC). The results showed that the containment strategies were effective and magnificently

suppressed the pandemic transmission. It was also found that softening personal protection too early might lead to the spread of disease. The SPSS modeler was also used to investigate the correlation between average daily temperatures and the growth rate of COVID-19 in infected countries [40]. It was shown that the pandemic rates were higher in case studies where the average temperature is lower. Finally, in [41] a coupled ordinary differential equation metapopulation model for different courses on the disease in different age groups were developed. It was shown that the economic lockdown could be safely reversed at any time without a substantial effect on the course of the disease. In addition, it was concluded that strict quarantines could not be necessary to keep the number of infected people low.

## 4 Artificial intelligence and COVID-19

Artificial intelligence (AI) has been used mostly for medical image segmentation and diagnosis to classify whether a patient has COVID-19 or what is the severity of the infection. The images used in these works were mostly from medical X-ray radiology or Computed Tomography (CT). Before presenting the AI methodologies used in COVID-19 detection and classification, a brief description of these medical imaging modalities is presented.

### 4.1 COVID-19 detection based on CT scan

X-ray radiology consists of beaming x-ray photons onto a part of body to be imaged and collecting the photons that pass through that part of the body. Depending on the body's tissue type, it will attenuate (block) some of the incident photons. This will create a shadowy image of the body on a detector located behind the body. X-ray radiology is used to examine bone structure and detect infections in the lungs. Computed tomography (CT) takes the ideas of X-ray radiography further by taking X-rays images of the body from multiple angles to produce cross-sectional images without dissecting the body. These cross-sectional images also called slices are tomographic images and these contain more detail medical information than the conventional x-rays radiography. CT images are used to detect abnormalities in the body like tumors and hemorrhage it can also be used to detect pulmonary embolisms, excess fluid, and pneumonia in the lungs [42, 43]. This makes it suitable for diagnosis of COVID-19 which is a disease that attacks the lungs and the respiratory system.

In their study, Pan Feng et al. seek to verify the change obtained in the chest images of patients with COVID-19 pneumonia. The study was carried out on 4-day intervals from the first day of diagnosis to the day of total recovery. Excluded from this study are patients with complicated

pneumonia with severe respiratory distress. For non-severe cases, the results of the chest scanner show a progress of lesions severity during the first 10 days, then stabilizes thereafter. According to this study, almost all the patients presented a spike of the disease around the 10th day, and the signs of improvement around the 14th day of the symptoms [44]. In a series of experiments carried out in 3 days on 51 patients, Yicheng Fang et al. studied the performance of 2 methods of medical examinations on patients with Covid-19. The results indicate that the sensitivity of chest CT to Covid-19 is higher than the RT-PCR technique (98% for CT versus 71% for PCR). When RT-PCR tests are negative, chest CT can therefore be used on patients with clinical and epidemiological characteristics of COVID-19, to confirm or refute the previous results [45]. Li Yan et al. also conducted the study to determine the rate of false diagnoses and the performance of CT scans on COVID-19. Their study was carried out on the first 51 patients confirmed by nucleic acid tests. The study confirmed the high performance of the chest CT which produced a low rate of false diagnosis on COVID-19 [46].

### 4.2 Image based (X-ray, CT) AI CoVID-19 detection and classification

The classification consists of separating images into groups. The three standards well known procedures to do that are supervised learning, unsupervised learning, and semi-supervised learning.

Supervised learning is an automatic task allowing a function to match input-output pairs [47]. The purpose of a supervised learning algorithm is to produce a function which maps the input-output (vector-supervision signal) pair. The algorithm will allow in an optimal scenario to correctly label the data to determine the classes. In the parallel world of human psychology, it is called conceptual learning [48]. Among the supervised learning algorithm used of the detection of COVID-19 are Convolutional Neural Network (CNN), Support Vector Machines (SVM), Logistic Regression (LR), Naive Bayes (NB), Linear Discriminant Analysis (LDA), Decision Trees (DT) and Random Forest (RF). Table 4 shows the summary of the classifications methods used in COVID-19 studies.

#### 4.2.1 Convolutional neural network (CNN)

The principle of Neural Network (NN) is based on the collection of nodes (called artificial neurons), which freely model neurons in the brain. Based on examples, without any prior knowledge, without being programmed, this system automatically generates identification characteristics. When the algorithm uses multiple layers of neurons it is known as Deep learning. A Convolutional Neural Network (CNN) is



a Deep Learning algorithm which takes an image as input, assign learnable weights to various features (objects) in the image so as to be able to differentiate one image from the other [9] [10].

Wang et al. in [49] used CNN with a dataset comprising of 13,800 chest x-ray radiography images from 13,725 patients so as to try and provide clinicians with a deeper insight into the critical factors affecting with COVID-19 cases. The reported an accuracy, sensitivity and positive prediction value (PPV) of 92.6%, 87.1% and 96.4% respectively. In [50], three models (ResNet50, InceptionV3 and InceptionResNetV2) based on CNN were proposed for detecting COVID-19 in pneumonia infected patients from chest X-ray radiography images. They used ROC analyses and confusion matrices to evaluate the performances of the three models and found that the ResNet50 model provided the best classification performance with an accuracy of 98%. In a retrospective and multi-center study carried out by Li et al. [51], CNN was employed for the detection of COVID-19. They extracted visual features from volumetric chest CT images of COVID patients and classified them. They reported that the method was not only able to detect COVID-19 case but also to distinguish it from other community acquired pneumonia and non-pneumonic lung diseases. In [52] a concept known as Transfer learning (where available data from one scenario is used to enhance accuracy of detection in a second scenario where there is lack of data) was used on X-ray images from patients with ordinary bacterial pneumonia, confirmed COVID-19 cases, and other normal infections. The goal of the work was to evaluate the performance of some state-of-the-art CNN architectures for medical image classification. They obtained an accuracy, sensitivity, and specificity of 96.78%, 98.66%, and 96.46% respectively and concluded that CNN with X-ray imaging might extract significant biomarkers related to COVID-19. Hemdan et al. [53] on their part implemented seven different CNN architectures with the aim of assisting radiologists in the automatic diagnoses of COVID-19 in X-ray images. They validated the architectures on 50 Chest X-ray images with half confirmed COVID-19 cases. They reported that the VGG19 and Dense Convolutional Network (DenseNet) models had the best performance both with an accuracy of of 90%.

#### 4.2.2 Support vector machines (SVM)

Support Vector Machines (SVM) are supervised learning methods used for regression, classification and also outlier detection. The aim of SVM is to find a hyperplane in an N-dimensional space (where N is the number of features) that markedly classifies the input data. In other words, SVM will work to find a plane that has the maximum distance between data points of separate classes. Support vectors

are those data points that are closest to the hyperplane. These data points affects the position and orientation of the hyperplane [67].

Barstugan et al. [54] presented an early detection of COVID-19 based on SVM. The algorithm was applied on abdominal Computed Tomography (CT) images. Four different image datasets of variable size (16x16, 32x32, 48x48, 64x64) were created from 150 CT images. Features were extracted through Grey Level Co-occurrence Matrix (GLCM), Local Directional Pattern (LDP), Grey Level Run Length Matrix (GLRLM), Grey-Level Size Zone Matrix (GLSZM), and Discrete Wavelet Transform (DWT) algorithms. SVM was then used to classify the extracted features. A maximum Sensitivity and accuracy of 97.56% and 98.71% respectively were obtained with 10-fold cross-validation and GLSZM feature extraction method. In [55] a combination of deep feature extractor and SVM was used to detect COVID-19 infection in X-ray images. The proposed model (combination of resnet50 and SVM) obtained an accuracy of 95.38%. In [57] SVM was used on features extracted from chest X-ray radiography images for early detection of COVID-19 cases. The features were extracted through a multi-level thresholding of the images. They obtained a classification accuracy of 98.82% on a total of 40 contrast-enhanced chest X-ray images.

Non-image data was also used with SVM and data from emergency care admission exams to detect COVID-19 cases. De Moraes et al. [56] used SVM and data from emergency care admission exams to detect COVID-19 cases. They collected data from 235 patients of which 43% were confirmed COVID-19 cases. They trained five machine learning algorithms namely logistic regression, random forests, gradient boosting trees, neural networks, and support vector machines on 70% of the patients, and evaluated their performance on the remaining 30%. They found out that the SVM had the best performance with an accuracy of 85% and concluded that the method could be used to target which patient needs a laboratory COVID-19 tests done on them.

#### 4.2.3 Logistic regression (LR)

In statistics, logistic regression is used to model the probability, each sample is assigned a probability between 0 and 1. It can be extended to model several classes of events in order to determine for example different objects in an image [68]. Although simpler than the CNN, logistic regression also could be applied in the in depth study of the manifestation of COVID-19. For instance, in [58] logistic regression was applied to values provided by ROC analysis in the aim of investigating clinical and CT features that indicates severity COVID-19. Through logistic regression analyses it was found that the clinical factors associated

with severe/critical COVID-19 pneumonia were patient older than 50 years, chest pain, dyspnea, comorbidities and cough among others. In [59] deep features from COVID-19 patient chest X-ray images were extracted using ResNet152 and then SMOTE was used to balance the data points of COVID-19 and Normal patients. Then finally, machine learning algorithms like Random Forest and XGBoost were used to classify according to the features. They obtained an accuracy of 97.3% for Random Forest and 97.7% for XGBoost.

#### 4.2.4 Naive Bayes (NB)

Naive Bayes classifiers are among the simplest Bayesian network models from the family of probabilistic classifiers. Coupled with the Kernel density estimation, they can reach high levels of precision in digital images classification [69]. In the study of COVID-19, it has also been used for classification. In [60], the authors combined conventional statistical and machine learning in order to extract features from CT images. The extracted features were then classified by hybrid classifier system based on Naive Bayes. Experimental evaluation of this method produced and accuracy of 96.07%.

#### 4.2.5 Linear discriminant analysis (LDA)

Linear discriminant analysis (LDA) is used to find a linear combination of features that characterizes or separates classes of objects or events in pattern recognition and machine learning. This resulting combination can be used as a linear classifier for dimensionality reduction before the final classification [70].

LDA was used in [61] with the aim of investigating the characteristics and rules of hematology changes in patients suffering from COVID-19. Clinical and laboratory test results of the patients were analyzed and different hematological parameters were fitted using LDA. The  $NLR \& RDW - SD$  combined parameter was found to be the best indicator of the severity of COVID-19 in patients with an accuracy of 93.8%.

#### 4.2.6 Decision trees (DT) and random forest (RF)

Decision trees is a technique that helps analyzing decisions by identifying the most likely strategy leading to the goal. Random Forest on its part is essentially a collection of Decision Trees whose results are accumulated into a final result. They have the ability to limit variance without increasing error due to bias. In medical practice, it is used to classify patient images [71]. In [62], the chest CT images of 176 patients with COVID-19 were used for

severity assessment. A random forest modeled and trained to evaluate the severity of COVID-19 in patients based on quantitative features. The RF model showed encouraging results with an accuracy of 87,5%. Shi et al. proposed an infection Size Aware Random Forest method (iSARF), their method had two steps, the first one consisted of categorizing different groups while the second classified the images [63]. They used an infection size feature defined as the ratio of the volume of infected regions to the total volume of whole segmented lung. This infection size was then used in a 3 level Random Forest classifier that classified it into 4 groups. They used a 5-fold cross-validation to evaluate the performance of the proposed algorithm and also compared it to other classifiers like logistic regression, support vector machine and neural network (NN). They obtained a sensitivity, specificity and accuracy of 90.7%, 83.3%, 87.9% respectively.

#### 4.2.7 U-net

U-Net was first proposed by Ronneberger et al. for segmentation of Biomedical images [72]. The U-Net architecture has two paths namely a contraction path or the encoder and an expanding path or the decoder. In the encoder, successive convolutional and max pool down-sampling layers are used to extract the context of an image while in the decoder the discriminative features learnt in the encoder are projected onto the pixel space (image) so as to obtain a semantically segmented image. The decoder is made up of a series of upsampling, concatenation and then convolution operation.

U-Net based algorithms were also used in the segmentation of medical images for the purpose of COVID-19 detection. Chen et al. proposed a new method called modified U-net structure to segment the regions of infected lungs with COVID-19. They used Aggregated Residual Network (ResNeXt) for learning and complex features from the original images. They also applied a soft attention mechanism that enhanced the model's ability to differentiate various symptoms of COVID-19 [64]. In [65], Attention U-Net was used with an adversarial critic model to improve its performance. They obtained an average dice score of 97.8% on 1047 chest X-ray images from three sources. In [66], two methods are proposed, namely the InfNet and the Semi-Inf-Net. The Inf-Net uses implicit Reverse Attention and explicit Edge Attention to ameliorate the detection of infected regions in CT lung images. The Semi-InfNet is a semi-supervised solution that helps to overcome the lack of high quality and labeled images. They carried out extensive experiments on COVID-19 datasets and showed that the proposed methods perform better than other segmentation methods.

#### 4.2.8 Unsupervised learning

Unsupervised learning have also been used in the study of COVID-19. Unsupervised learning, unlike supervised learning, searches for previously undetected prototypes in a data stream without pre-existing labels and minimum human intervention. It makes it possible to model the densities of probability on the entries. This algorithm makes it possible to detect abnormal parts of data which do not correspond to any group, its application is in the field of density estimation in statistics [10]. Among the unsupervised learning used in COVID-19 is k-means clustering which is a vector quantization algorithm. It partitions  $n$  observations into  $k$  clusters in which each annotation belongs to the cluster with the neighboring mean, serving as the princeps of the cluster [73].

### 5 Datasets of COVID-19

In both mathematical modeling and AI, data is the raw material. So the first step in the development of COVID-19 applications is data collection. Over the course of few months there are multiple datasets that have been put online in regards to the COVID-19. Most if not all of these datasets are open source meaning that they are free for anyone to download and use. Also, they are constantly being updated with new data from the field. Table 4 presents a collection of the open source datasets explored. The following presents a comprehensive description of these datasets.

Dong et al. [74] currently provides one of the most complete database of the COVID-19 situation. The database known as the 2019 Novel Coronavirus Visual Dashboard operated is maintained by the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE). They obtained data from about 18 sources such as the WHO, CDC and other governments agencies, compiled and shared them in the form of an interactive map of the COVID-19 situation map. The database includes number of daily contamination, active, recovery and death. It also contains the location (state/province, country, longitude, latitude), number of people tested, incident rate and hospitalization rate. Xu et al. [75] are currently collecting and sharing health information on persons with COVID-19 from local to national level, together with other information from online reports. The data are localized geographically and also indicate aspects like the symptoms and the dates of confirmation and admission and also travel record.

Cohen et al. [76] created a COVID-19 Image Database by collecting X-ray images from various websites as well as publications. The database is made up of 345 X-ray images. Zhao et al. [77] created a computed tomography (CT) image database currently May 01, 2020) containing 349 images

of confirmed COVID-19 cases along with 398 images of non-COVID-19 cases. The CT images are gathered from several COVID19-related papers. Ma et al. provided a dataset containing 20 labeled COVID-19 CT images of the left lung, the right lung and the infection type. The labeling was done by two radiologists [78]. The aim of their work was to establish a benchmark for CT image segmentation of lungs in regards to COVID-19. Two radiologists in based in Oslo, Norway have shared two CT datasets, the COVID-19 CT segmentation dataset (with 100 axial CT slices) and the Segmentation dataset nr. 2 (with 829 CT slices) from more than 60 patients [79]. The databases were manually segmented by radiology experts.

Chen et al. shared a COVID-19 twitter dataset [80]. This dataset contains an ongoing collection of tweets IDs associated with COVID-19 and which started from January 28, 2020. Such tweet IDs include “COVID-19”, “Coronavirus”, “Pandemic” and so on. They also tracked certain accounts like that of the WHO, CDCgov and HHSgov. As of May 01, 2020, they collected more than 10 million tweets in many languages. Rabindra [81] is also collecting tweets using the LSTM model deployed on a website. The model continuously monitors real-time Twitter feed for COVID-19 related tweets. It uses filters such as language “en”, and tweeter keywords like “covid19”, “coronavirus”, “covid” and so on. As of May 01, 2020, more than 30million tweets were collected.

*Havard Dataverse* also provides the Global News dataset which contains COVID-19 related *Global English news* from *GDELT* [82] and the climate dataset which contains time series temperature, humidity, air quality and other monitored data in China from January 1, 2020 [83]. *The Coronacases Initiative* which is a pro bono initiative of *RAIOSS Desenvolvimento Ltda* and *Livon Saúde Ltda*, also provides information on COVID-19 cases on their website [84].

In COVID-19 and other pandemic studies other datasets such as population density, mobility, Security incidents, economic situation, humanitarian condition data, and healthcare workforce are important data that will ensure the accuracy of the studies. Several sources provide those datasets. One of such sources is *The WorldPop* which shares spatial demographic datasets from Africa, Asia and central and South America [85]. Some of the datasets provided by WorldPop are population data, births, internal migration, age and sex data, administrative areas and global flight data. *The Humanitarian Data Exchange (HDX)* coordinated by the *UN Office for the Coordination of Humanitarian Affairs (OCHA)* shares more than 17,000 humanitarian datasets form 253 locations around the globe [86]. The WHO on its part shares the *Global Health Workforce Statistics* [87]. The dataset includes data on the number of health workers as well as hospital bed capacity in each country. The



**Table 4** A collection of the open source dataset sources and their links

No.	Dataset name	Data type	Size <sup>†</sup>	Link
1	COVID-19	Text, values(Prevalence)	–	<a href="https://github.com/CSSEGISandData/COVID-19">https://github.com/CSSEGISandData/COVID-19</a>
2	n-CoV 2019	Text, numbers (Prevalence)	–	<a href="https://github.com/beoutbreakprepared/nCoV2019~~">https://github.com/beoutbreakprepared/nCoV2019~~</a>
3	COVID-19 image data collection	chest X-ray or CT images	345	<a href="https://github.com/ieee8023/covid-chestxray-dataset">https://github.com/ieee8023/covid-chestxray-dataset</a>
4	COVID-CT-Dataset: A CT Scan Dataset about COVID-19	CT images	398	<a href="https://github.com/UCSD-AI4H/COVID-CT">https://github.com/UCSD-AI4H/COVID-CT</a>
5	COVID-19-CT-Seg-Benchmark	CT images of Lungs	525	<a href="https://gitee.com/junma11/COVID-19-CT-Seg-Benchmark">https://gitee.com/junma11/COVID-19-CT-Seg-Benchmark</a>
6	COVID-19 CT t segmentation dataset	CT images	100	<a href="http://medicalsegmentation.com/covid19/">http://medicalsegmentation.com/covid19/</a>
7	Segmentation dataset nr. 2	CT images	829	<a href="http://medicalsegmentation.com/covid19/">http://medicalsegmentation.com/covid19/</a>
8	COVID-19-TweetIDs	Text (Social media)	100 million	<a href="https://github.com/echen102/COVID-19-TweetIDs">https://github.com/echen102/COVID-19-TweetIDs</a>
9	CORONA VIRUS (COVID-19)	Text (Social media)	30 million	DOI: 10.21227/781w-ef42
10	TWEETS DATASET	Text (News)	–	<a href="https://doi.org/10.7910/DVNTU0JDP">https://doi.org/10.7910/DVNTU0JDP</a>
11	China datalab “Global News”	Values (Climatic data)	–	<a href="https://doi.org/10.7910/DVNTU0JDP">https://doi.org/10.7910/DVNTU0JDP</a>
12	China datalab “Climate”	3D CT images	10	<a href="https://coronacases.org">https://coronacases.org</a>
13	Coronacases Initiative	Values (Demography)	–	<a href="https://www.worldpop.org">https://www.worldpop.org</a>
14	WorldPop	Text, values (Humanitarian)	18,064	<a href="https://data.humdata.org/dataset">https://data.humdata.org/dataset</a>
15	HDX	Values (Health workforce)	–	<a href="https://www.who.int/hrh/resources/en/">https://www.who.int/hrh/resources/en/</a>
16	WHO Global Health Workforce Statistics	Values (Mobility data)	–	<a href="https://www.apple.com/covid19/mobility">https://www.apple.com/covid19/mobility</a>
17	Apple Mobility Trends Report	Values (Mobility data)	–	<a href="https://www.google.com/covid19/mobility">https://www.google.com/covid19/mobility</a>
18	Google COVID-19 Community Mobility Reports	Values (COVID-19 testing data)	–	<a href="https://www.acaps.org/covid19-government-measures-dataset">https://www.acaps.org/covid19-government-measures-dataset</a>
19	Our World in Data	Text and Values (Management measures)	–	<a href="https://www.acaps.org/covid19-government-measures-dataset">https://www.acaps.org/covid19-government-measures-dataset</a>
20	ACAPS	Values (Security incidents)	–	<a href="https://www.acleddata.com">https://www.acleddata.com</a>
21	The Armed Conflict Location & Event Data Project (ACLED)	Values (Economic outlook)	–	<a href="https://www.imf.org/en/Topics/imf-and-covid19/Policy-Responses-to-COVID-19">https://www.imf.org/en/Topics/imf-and-covid19/Policy-Responses-to-COVID-19</a>
22	The International Monetary Fund (IMF)	Values (Economic outlook)	–	<a href="https://ourworldindata.org/covid-testing">https://ourworldindata.org/covid-testing</a>
23	BFA Global	Various	–	<a href="https://c3.ai/products/c3-ai-covid-19-data-lake/">https://c3.ai/products/c3-ai-covid-19-data-lake/</a>
24	C3.ai COVID-19 Data Lake	Data, literature	–	<a href="https://github.com/HzFu/COVID19_imaging_AI_paper_list">https://github.com/HzFu/COVID19_imaging_AI_paper_list</a>
24	COVID-19 Imaging-based AI Research Collection			

<sup>†</sup>The Dataset sizes are all as of May 01, 2020

tech-giants *Apple* and *Google* both released mobility reports on COVID-19. Apple called their dataset *Mobility Trends Reports* [88] while *Google* called it *Google COVID-19 Community Mobility Reports* [89]. Both presents aggregated data that registers the daily use of various modes of transportation (walking, driving, transit) since the start of February 2020 as well as places visited or stayed in by users of their services. The data was collected from customer requests for directions or location in Apple and in Google Maps. They also offer a useful visualization tool of the data. *Our World in Data* on its part provides COVID-19 Testing dataset where they collect data that are based on tests carried out to establish if a person is currently infected [90]. *ACAPS* [91] provides a dataset of Government Measures Dataset also provides Government Measures implemented by Governments all around the world in response to COVID-19 while *The Armed Conflict Location & Event Data Project (ACLED)* [92] provides security incidents related to COVID19 dataset. *The International Monetary Fund (IMF)* [93] and *BFA Global* [94] both provide datasets on the key economic responses of governments and the effect of COVID19 management measures on economy.

Lastly, the software provider *C3.ai* compiled, cleaned, structured and standardized COVID-19 data from most of the sources presented in this paper [95]. The initiative known as *C3.ai COVID-19 Data Lake* contains analysis-ready COVID-19 data in one place. The service is free and the datasets are updated continuously. It contains everything from time-series data to case reports. Also, a github repository was created to collect COVID-19 images regarding AI research papers and datasets. It contain 19 datasets, 11 review papers, 18 clinical papers on Covid19 images, 54 AI-related papers, 54 articles on CXR methods, and 1 paper on Line Artefact Quantification in Lung Ultrasound Images [96].

## 6 Discussion and conclusion

The use of mathematical modeling and AI with COVID-19 data will increase our knowledge on the disease propagation evaluating prevention measures as well as early and accurate detection of the disease in patients. However, to arrive at this end a lot of data is needed to explore various models and AI algorithms. The data available up till now are mostly of medical images (for diagnosis) and text based data (for social impact analysis). While the later may be generated by and readily available to a large number people, the former on the other hand can only be generated in a specialized institution by a specialized professional. This means that data in low resource setting are not available as these places do not have the sophisticated imaging equipment needed to

generate such images [97]. Also it is well known in data science that datasets from different geographical locations may not hold the same information and this is especially true in terms of healthcare data. More data types are therefore needed that can be easily generated easily anywhere on the Globe so as to enhance and render the application of the mathematical models and AI algorithm possible for many. These data types could be physiological measurements such as ECG, SPO2, body temperature that could be obtained using wearable devices [98]. Data concerning the type of preventive measures implemented by authorities are also not well documented. In this work only a few of the dataset found provided that information. However, this information could help in the examination and optimization of the set measures thereby improving the situation.

In mathematical modeling, most of the articles found in the writing of this paper are of COVID-19 dynamics. However, modeling can be done with appropriate datasets to explore the effect of the variables like climate and preventive measure on the spread of COVID-19 as explained earlier. There is also not many studies on the correlation of environmental and climatic conditions to the COVID-19 propagation in the work only two articles were found that addresses this issue and they both provide in different and interesting way of looking at the propagation of this diseases [27, 28]. Simulation of second and third waves of COVID-19 outbreaks will also help to enhance surveillance. As countries start easing social restriction measures, a study is needed to estimate possible hotspots for new outbreaks.

AI (deep learning) is powerful tool for early and accurate diagnosis of COVID-19 and many articles have addressed it. Most of them apply convolutional neural networks (CNN) in their work for medical image classification. Few other studies apply the Random forest and Support Vector Machines. There are also some that applied U-Net and its variations for the segmentation of CT and X-Ray images. The authors of the AI algorithms reviewed here all claimed that their algorithm performs very well on test data. However, it is well known that good performance of an algorithm on test data does not mean that it will perform similarly when deployed on the field. This is due to fact that in real life the data is more prone to noise and other artefact that are not usually present in the training and test data. The lack of diverse annotated images is also not helping the situation. In this review only 2 out of 18 studies were found to used annotated data from radiologists. Collaboration is needed between clinicians and AI experts in other to build a huge amount of annotated images of COVID-19. Also human in the loop or human augmentation can be another solution to overcome the problem caused by the disparity of an algorithm's performance when applied to test data and when applied in the real world. Most of the

studies reviewed used existing models while a few used well known models with some modifications. Those used with some modifications performed slightly better than the others stressing the need of developing hybrid models to build better and robust architectures. Much work is also needs to be done in terms of drug and/or vaccine discovery, treatment selection and contamination risk assessment for medical personnel [99]. Finally, since most of the AI research objective on COVID-19 is to find the optimal solution for diagnosis, other algorithms like Genetic Programming and Boosting (AdaBoost) should be explored so as to clear any doubt regarding their performances.

In conclusion, COVID-19 has spread rapidly all over the world creating an emergency situation. Mathematical modeling and AI have both shown to be reliable tools in the fight against this pandemic. Most of the modeling done were based on the Susceptible-Exposed-Infected-Removed (SEIR) model and the Susceptible-infected-recovered (SIR) model while most of the AI implementations were Convolutional Neural Network (CNN) on X-ray and CT images. Several datasets concerning the COVID-19 have been collected and shared open source. However, much work is needed to be done in terms of providing the public with a wide variety of data types and from many regions as possible. Also, other AI and modeling applications in healthcare should be explored in regards to this COVID-19.

**Acknowledgements** Not applicable.

## Compliance with Ethical Standards

**Conflict of interests** The authors declare that they have no conflict of interest.

## References

1. WHO Emergencies Coronavirus Emergency Committee Second Meeting. [https://www.who.int/docs/default-source/coronaviruse/transcripts/ihr-emergency-committee-for-pneumonia-due-to-the-novel-coronavirus-2019-ncov-press-briefing-transcript-30012020.pdf?sfvrsn=c9463ac1\\_2](https://www.who.int/docs/default-source/coronaviruse/transcripts/ihr-emergency-committee-for-pneumonia-due-to-the-novel-coronavirus-2019-ncov-press-briefing-transcript-30012020.pdf?sfvrsn=c9463ac1_2). Accessed on May 01, 2020
2. Sohrabi C, et al. (2020) World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). In: International Journal of Surgery
3. World Health Organization Novel Coronavirus (2019-nCoV): situation report, 10. [https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200130-sitrep-10-ncov.pdf?sfvrsn=d0b2e480\\_2](https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200130-sitrep-10-ncov.pdf?sfvrsn=d0b2e480_2). Accessed on May 01, 2020
4. World Health Organization Novel Coronavirus (2019-nCoV): situation report, 10. [https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200421-sitrep-92-covid-19.pdf?sfvrsn=38e6b06d\\_4](https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200421-sitrep-92-covid-19.pdf?sfvrsn=38e6b06d_4). Accessed on May 01, 2020
5. Chen N et al (2020) Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. The Lancet 395(10223):507–513
6. Nokes DJ, Anderson RM (1988) The use of mathematical models in the epidemiological study of infectious diseases and in the design of mass immunization programmes. Epidemiology & Infection 101(1):1–20
7. Kruijschaar ME, Barendregt JJ, Hoeymans N (2002) The use of models in the estimation of disease epidemiology. Bull World Health Org 80:622–628
8. Wynants L et al (2020) Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. In: bmj 369
9. Ng A (2016) What artificial intelligence can and can't do right now. In: Harvard Business Review 9
10. Ng A (2017) Artificial intelligence is the new electricity. In: Presentation at the Stanford MSx Future Forum
11. Yu K-H, Beam AL, Kohane IS (2018) Artificial intelligence in healthcare. Nat Biomed Eng 2(10):719–731
12. Russel S, Norvig P et al (2013) Artificial intelligence: A modern approach. Pearson Education Limited
13. Shi F et al (2020) Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19. arXiv:2004.02731
14. Diekmann O, Heesterbeek H, Britton T (2012) Mathematical tools for understanding infectious disease dynamics, vol 7. Princeton University Press, Princeton
15. Pastor-Satorras R, Vespignani A (2001) Epidemic spreading in scale-free networks. Phys Rev Lett 86(14):3200
16. Boguná M, Pastor-Satorras R, Vespignani A (2003) Absence of epidemic threshold in scale-free networks with degree correlations. Phys Rev Lett 90(2):028701
17. Small M, Tse CK, Walker DM (2006) Super-spreaders and the rate of transmission of the SARS virus. Physica D: Nonlinear Phenomena 215(2):146–158
18. Small M, Tse CK (2005) Small world and scale free model of transmission of SARS. Int J Bifurcation Chaos 15(05):1745–1755
19. Small M, Tse CK (2005) Clustering model for transmission of the SARS virus: application to epidemic control and risk assessment. Physica A: Statistical Mechanics and its Applications 351(2-4):499–511
20. Wang Y et al (2003) Epidemic spreading in real networks: An eigenvalue viewpoint. In: 22nd international symposium on reliable distributed systems, 2003. Proceedings. IEEE, pp 25–34
21. Chakrabarti D et al (2008) Epidemic thresholds in real networks. ACM Trans Inf Syst Secur (TISSEC) 10(4):1–26
22. Gross T, D'Lima CJD, Blasius B (2006) Epidemic dynamics on an adaptive network. Phys Rev Lett 96(20):208701
23. Zhan C et al (2020) Modelling and prediction of the 2019 Coronavirus Disease spreading in China incorporating human migration data. In: Available at SSRN 3546051
24. Radulescu A, Cavanagh K (2020) Management strategies in a SEIR model of COVID 19 community spread. arXiv:2003.11150
25. Gupta R et al (2020) SEIR and Regression Model based COVID-19 outbreak predictions in India. In: medRxiv
26. Prem K et al (2020) The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: A modelling study. In: The Lancet Public Health
27. Zhang H, Guo X, Zeng Y (2020) Transmissibility of COVID-19 and its association with temperature and humidity
28. Eikenberry SE et al (2020) To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. In: Infectious disease modelling
29. Liu T-H (2020) A Time-dependent SIR model for COVID-19 with Undetectable Infected Persons
30. Song PX et al (2020) An epidemiological forecast model and software assessing interventions on COVID-19 epidemic in China. In: medRxiv
31. Biswas K, Khaleque A, Sen P (2020) Covid-19 spread: Reproduction of data and prediction using a SIR model on Euclidean network. arXiv:–2003

32. Chikina M, Pegden W (2020) Modeling strict age-targeted mitigation strategies for COVID-19. arXiv:2004.04144
33. Singh R, Adhikari R (2020) Age-structured impact of social distancing on the COVID-19 epidemic in India. arXiv:2003.12055
34. González R (2020) Different scenarios in the Dynamics of SARS-Cov-2 Infection: An adapted ODE model. arXiv:2004.01295
35. Crokidakis N (2020) Data analysis and modeling of the evolution of COVID-19 in Brazil. arXiv:2003.12150
36. Vyasarayani CP, Chatterjee A (2020) New approximations, and policy implications, from a delayed dynamic model of a fast pandemic. arXiv:2004.03878
37. Chen T-M et al (2020) A mathematical model for simulating the phase-based transmissibility of a novel coronavirus. *Infectious diseases of poverty* 9(1):1–8
38. Zhao Z et al (2020) A mathematical model for estimating the age-specific transmissibility of a novel coronavirus. In: medRxiv
39. Wan H, Cui J-A, Yang G-J (2020) Risk estimation and prediction by modeling the transmission of the novel coronavirus (COVID-19) in mainland China excluding Hubei province. In: medRxiv
40. Pirouz B et al (2020) Relationship between Average Daily Temperature and Average Cumulative Daily Rate of Confirmed Cases of COVID-19. In: medRxiv
41. Pérez-García VM Relaxing quarantine after an epidemic: A mathematical study of the Spanish COVID-19 case
42. Prince JL, Links JM (2006) Medical imaging signals and systems. Pearson Prentice Hall, Upper Saddle River
43. Brant WE, Helms CA (2012) Fundamentals of diagnostic radiology. Lippincott Williams & Wilkins
44. Pan F et al (2020) Time course of lung changes on chest CT during recovery from 2019 novel coronavirus (COVID-19) pneumonia. In: *Radiology*, pp 200370
45. Fang Y et al (2020) Sensitivity of chest CT for COVID-19: comparison to RT-PCR. In: *Radiology*, pp 200432
46. Li Y, Xia L (2020) Coronavirus disease 2019 (COVID-19): role of chest CT in diagnosis and management. *Am J Roentgenol* 214:1–7
47. Russell SJ, Norvig P (2016) Artificial intelligence: A modern approach. Malaysia
48. Mohri M, Rostamizadeh A, Talwalkar A (2018) Foundations of machine learning. MIT Press, Boca Raton
49. Wang L, Wong A (2020) COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest radiography images. arXiv:2003.09871
50. Narin A, Kaya C, Pamuk Z (2020) Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. arXiv:2003.10849
51. Li L et al (2020) Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct. In: *Radiology*, pp 200905
52. Apostolopoulos ID, Mpesiana TA (2020) Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. In: *Physical and Engineering Sciences in Medicine*, pp 1
53. Hemdan EE-D, Shouman MA, Karar ME (2020) Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images. arXiv:2003.11055
54. Barstugan M, Ozkaya U, Ozturk S (2020) Coronavirus (covid-19) classification using ct images by machine learning methods. arXiv:2003.09424
55. Sethy PK, Behera SK (2020) Detection of coronavirus disease (covid-19) based on deep features. In: Preprints 2020030300, pp 2020
56. de Moraes Batista AF et al (2020) COVID-19 diagnosis prediction in emergency care patients: a machine learning approach. In: medRxiv
57. Hassanien AE et al (2020) Automatic x-ray COVID-19 lung image classification system based on multi-level thresholding and support vector machine. In: medRxiv
58. Li K et al (2020) The clinical and chest CT features associated with severe and critical COVID-19 pneumonia. In: *Investigative radiology*
59. Kumar R (2020) Accurate prediction of COVID-19 using chest x-ray images through deep feature learning model with smote and machine learning classifiers. In: medRxiv
60. Farid AA, Selim GI, Khater HAA (2020) A novel approach of CT images feature analysis and prediction to screen for corona virus disease (COVID-19). *Int J Sci Eng Res* 11(3):1–9
61. Li C et al (2020) Preliminary study to identify severe from moderate cases of COVID-19 using NLR&RDW-SD combination parameter. In: medRxiv
62. Tang Z et al (2020) Severity assessment of coronavirus disease 2019 (COVID-19) using quantitative features from chest CT images. arXiv:2003.11988
63. Shi F et al (2020) Large-scale screening of covid-19 from community acquired pneumonia using infection size-aware classification. arXiv:2003.09860
64. Chen X, Yao L, Zhang Y (2020) Residual attention U-net for automated multi-class segmentation of COVID-19 chest CT images. arXiv:2004.05645
65. Gaál G, Maga B, Lukács A (2020) Attention U-Net Based Adversarial Architectures for Chest X-ray Lung Segmentation. arXiv:2003.10304
66. Fan D-P et al (2020) Inf-Net: Automatic COVID-19 Lung Infection Segmentation from CT Scans. arXiv:2004.14133
67. Vapnik Vladimir (2005) Universal learning technology: Support vector machines. *NEC J Adv Technol* 2(2):137–144
68. Hosmer DW Jr, Lemeshow S, Sturdivant RX (2013) Applied logistic regression, vol 398. Wiley, New York
69. Mayo M, Frank E (2020) Improving naive Bayes for regression with optimized artificial surrogate data. *Appl Artif Intell* 34(6):484–514
70. Hastie T, Tibshirani R (1996) Discriminant adaptive nearest neighbor classification and regression. In: *Advances in neural information processing systems*, pp 409–415
71. Rutkowski L et al (2014) A new method for data stream mining based on the misclassification error. *IEEE Trans Neural Netw Learn Syst* 26(5):1048–1059
72. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: *International conference on medical image computing and computer-assisted intervention*. Springer, Berlin, pp 234–241
73. Xia C et al (2020) Distributed K-Means clustering guaranteeing local differential privacy. *Comput Secur* 90:101699
74. Dong E, Du H, Gardner L (2020) An interactive web-based dashboard to track COVID-19 in real time. In: *The Lancet infectious diseases*
75. Xu B et al (2020) Epidemiological data from the COVID-19 outbreak, real-time case information, vol 7
76. Cohen JP, Morrison P, Dao L (2020) COVID-19 image data collection. arXiv:2003.11597
77. Zhao J et al (2020) Covid-ct-dataset: A ct scan dataset about covid-19. arXiv:2003.13865
78. Jun M et al (2020) OVID-19 CT Lung and Infection Segmentation Dataset (Version Version 1.0) [Data set]. In: Zenodo. <https://doi.org/10.5281/zenodo.3757476>
79. COVID-19 CT segmentation datasets. <http://medicalsegmentation.com/covid19/>. Accessed May 01, 2020
80. Chen E, Lerman K, Ferrara E (2020) Covid-19: The first public coronavirus twitter dataset. arXiv:2003.07372



81. Lamsal R (2020) Coronavirus (COVID-19) Tweets Dataset. In: IEEE-dataport. <https://doi.org/10.21227/781w-ef42>
82. China Data Lab Global News dataset. UNF:6:JTLzVY820iwN4knGfYaFiA==\[fileUNF]. Accessed on May 01, 2020
83. China Data Lab Climate dataset. <https://doi.org/10.7910/DVN/XETLSS>. Accessed on May 01, 2020
84. Coronacases dataset. <https://coronacases.org/forum/coronacases-org-helping-radiologists-to-help-people-in-more-than-100-countries-1>. Accessed on May 01, 2020
85. WorldPopdataset. <https://www.worldpop.org>. Accessed on May 01, 2020
86. The Humanitarian Data Exchange (HDX). <https://data.humdata.org/dataset>. Accessed on May 01, 2020
87. The 2018 update, Global Health Workforce Statistics, World Health Organization, Geneva. <https://www.who.int/hrh/resources/en/>. Accessed on May 01, 2020
88. Apple Mobility Trends Reportst. <https://www.apple.com/covid19/mobility>. Accessed on May 01, 2020
89. Google COVID-19 Community Mobility Reports. <https://www.google.com/covid19/mobility>. Accessed on May 01, 2020
90. World in Data COVID-19 Testing dataset. <https://www.acaps.org/covid19-government-measures-dataset>. Accessed on May 12, 2020
91. ACAPS. COVID19 Government Measures Dataset. <https://www.acaps.org/covid19-government-measures-dataset>. Accessed on May 12, 2020
92. Armed Conflict Location &Event Data Project (ACLED). <https://www.acleddata.com>. Accessed on May 12, 2020
93. IMF. Policy Responses to Covid-19. <https://www.imf.org/en/Topics/imf-and-covid19/Policy-Responses-to-COVID-19>. Accessed on May 12, 2020
94. BFA Global. Impact of COVID-19 on Financial Lives in Eight Countries. <https://ourworldindata.org/covid-testing>. Accessed on May 12, 2020
95. C3.ai COVID-19 Data Lake. <https://c3.ai/products/c3-ai-covid-19-data-lake/>. Accessed on May 01, 2020
96. COVID-19 Imaging-based AI Research Collection. [https://github.com/HzFu/COVID19\\_imaging\\_AI\\_paper\\_list](https://github.com/HzFu/COVID19_imaging_AI_paper_list). Accessed on May 12, 2020
97. Schlecht LLA (2010) Product development of a device for manufacturing medical equipment for use in low-resource settings
98. Tobin MJ (2020) Basing Respiratory Management of Coronavirus on Physiological Principles. In: American Journal of Respiratory and Critical Care Medicine, (2020)
99. McCall B (2020) COVID-19 and artificial intelligence: protecting health-care workers and curbing the spread. *The Lancet Digital Health* 2(4):e166–e167

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Youssoufa Mohamadou** obtained his BSc in Electrical and Electronics Engineering from the Islamique University of Technology in Dhaka, Bangladesh. He obtained his combined Masters and PhD in Biomedical Engineering from Kyoung hee University in South Korea. Currently, he is a Senior Lecturer and researcher at the University of Ngaoundere and at Université des Montagnes both in Cameroon. He works include signal processing, Artificial

intelligence and the design and development of medical devices adapted for low income communities.



**Aminou Halidou** is a Cameroonian PhD holder since 2014 from Huazhong University of Science and Technology (HUST) in Wuhan, China. His PhD research concerned on-road accident prevention based on pedestrian detection technology and he is thus an expert on digital image processing and computer vision. He holds 6 years' experience as a Lecturer at the University of Yaounde in Cameroon, currently he is the Head of the

Computer Science Department.



**Pascaline Tiam Kapen** is a Senior Lecturer at the University Institute of Technology Fotso-Victor of Bandjoun, of the University of Dschang, Cameroon. He received his Ph.D degree in Physics, option : Mechanics-Energetics at the University of Dschang, Cameroon in 2016. He is currently the Secretary General of the thematic research program in energy of the African and Malagasy council of Higher Education. His research interests are Computational Fluid

Dynamics, Aerodynamics, Riemann's solvers, Renewable Energy, Mathematical Modeling, Biomedical Engineering, Prototyping of low-cost and highly performing medical devices.