A Systematic Review on the Identification and Diagnosis of Clinical Characteristics of COVID-19 Patients

Poonam Phogat[a], Rajat Chaudhary[b]

[a]Computer Science & Engineering, SGT University, Gurugram, Haryana (India)
[b]Computer Science & Engineering, Bharat Institute of Engineering & Technology, Hyderabad, Telangana (India)

Abstract

World Health Organization (WHO) acknowledged the coronavirus disease 2019 (COVID-19) as one of the global diseases. The novel coronavirus disease is named Severe Acute Respiratory Syndrome Coronavirus (SARS-COV2) is a mild to severe respiratory disease having fever, cough, and shortness of breath as initial symptoms. This virus disseminates through contact with infectious persons, touching contaminated surfaces and infectious air droplets. The virus invades into healthy cells of the body especially the lungs causing the respiratory problems and sometimes causes organ failure by killing healthy cells which eventually leads to death. The origin of coronavirus disease was zoonotic, as the initial cases had been reported from animals in Wuhan. This paper presents the literature review on the identification and diagnosis of clinical characteristics of COVID-19 patients. Furthermore, the treatment and remedies of COVID-19 patients are discussed utilising machine learning and prediction models. The datasets related to the cases of COVID-19 patients are also discussed in this paper. Finally, the research gaps, open issues, and future directions are discussed.

Keywords

COVID-19, Diagnosis, Prediction Models

1. Introduction

COVID-19 is an epidemic virus that was initially spread within China and later it was declared by the WHO as a pandemic disease that was spread globally. On 31st December 2019, the first case of the threatening disease in human beings was disclosed in Wuhan; a city of China. SARS-COV-2 is the name given to this coronavirus disease which is mild to serious disease with the initiative symptoms retaining cough, fever and difficulty in breathing. This virus disseminates through contact with infectious persons, touching contaminated surfaces and infectious air droplets. The virus invades into healthy cells of the body especially the lungs causing the respiratory problems and sometimes causes organ failure by killing healthy cells which eventually leads to death. The origin of coronavirus disease was zoonotic, as the initial cases had been reported from animals in Wuhan [1]. The foremost epidemic named Severe Acute Respiratory Syndrome (SARS-COVID-19) was reported in 2002-2003.

1.1. Data Statistics on COVID-19

Figure 1 shows the WHO statistics on COVID-19 related to the total number of active cases, cured cases, and deaths cases [1]. The statistical reports are summarized based on state-wise cases in India, month-wise reports of continents, and the total number of confirmed cases by the WHO region. The first human case of this viral disease was spotted in Southern China in Nov. 2002. After this second epidemic viral disease occurred in 2012 named Middle East Respiratory Syndrome (MERS) in Saudi Arabia. The current outbreak of viral disease is reported in Dec. 2019 in China which is named SARS-COVID-19 [2]. As per the WHO report on COVID-19, a total of 26,47,663 active cases, 50,921 death cases, and 19,19,842 cured cases reported in India till July 31st, 2020. Maharashtra is ranking highest with 5,95,865 active cases, 20,037 death cases, and 4,17,123 cured cases followed by Tamil Nadu having 3,38,055 active cases, 5,766 death cases, and 2,78,270 cured cases.

The state-wise data shown in statistics are taken from the WHO report. As of January 6th, 2020, 44 confirmed cases were reported in the Americas region. According to WHO month-wise report from January to July COVID-19 confirmed cases are becoming exponentially increased, dated July 31st, 2020, Americas region was on top with 1,15,61,554 confirmed cases followed by Europe with 37,79,672 confirmed cases. Western Pacific region was ranking last having 4,14,606 confirmed cases. Thus, a rapid spread of these viral diseases was noticed due to its human to human transmission till Feb 9, 2020 with 37,289 confirmed cases ,which included 813 deaths in china and 302 confirmed cases in multiple countries.

ISIC'21: International Semantic Intelligence Conference, February 25-27, 2021, Delhi, India
poonamphogat07@gmail.com (P. Phogat); rajat@biet.ac.in (R. Chaudhary)
0000-0002-6554-918X (R. Chaudhary)

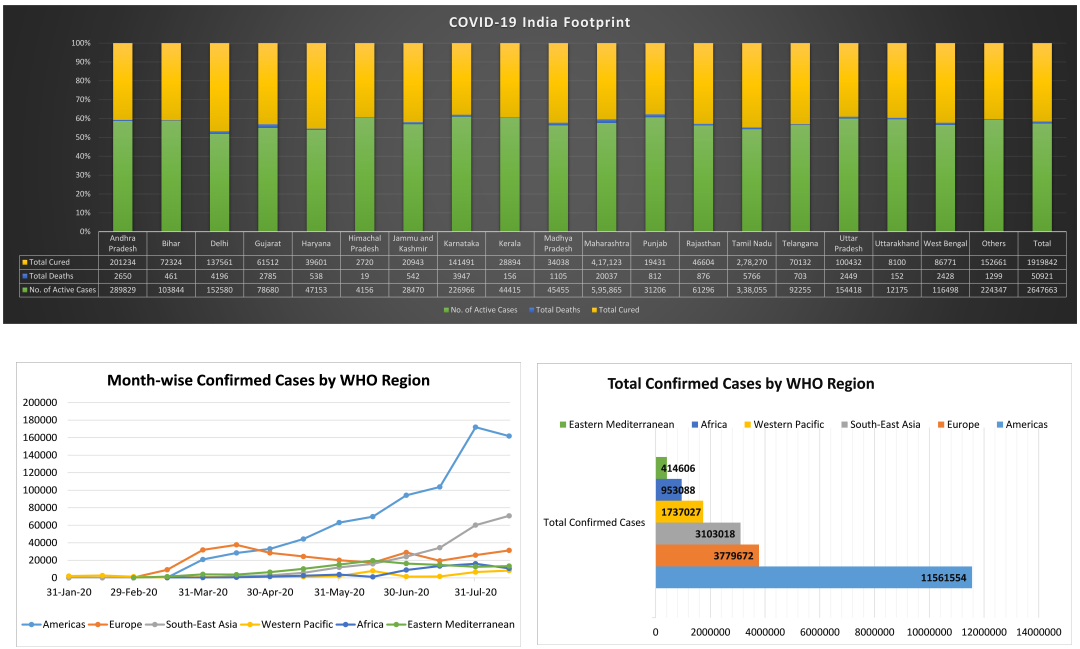

Figure 1: WHO statistics on COVID-19.

1.2. Transmission Stages of COVID-19 Infection

WHO classified novel coronavirus into four transmission stages as presented in Figure 2.

- *Stage 1:* This stage begins as the initial spread of the disease with the travel history of a person from infected areas. Such people were kept quarantined for a time period of 15 days in hospitals. The infection would be controllable at this stage because of the easy tracing of infection sources.
- *Stage 2:* It is the local transmission in which cases were reported in the peoples who came into contact with infected persons, hence till this stage, sources were traceable. Lockdown became necessary to prevent further spread of this viral.
- *Stage 3:* It is community transmission having neither travel history nor in contact with infected persons. Sources became untraceable due to its widespread. The reason behind its transmission was the presence of air droplets due to exhalation by the infected person and inhalation of those droplets by a normal person. Some necessary steps were taken to control the disease by applying social distancing, masks, and sanitizers. Let us assume that the model yields the droplet lifetime is ω, hence the infection rate constant is calculated as follows: $r^1 = f(\omega), r^2 = g(r^1)$, where f, g are the functions.
- *Stage 4:* It is an uncontrollable stage. This is the worst stage as the transmission becomes pervasive and the number of cases crossed the threshold limit. China was reported as the first union WHO faced all these four stages [3]. The reaction based pandemic model equation is calculated on the basis of infection growth rate as $exp(r^2 \times t)$, where t is the time interval during the diagnosis phase.

To break the chain of viral transmission, PM Modi announced 'JANTA CURFEW' on March 22nd, 2020 in India. The global situation was unbalancing very dangerously at that time due to the extremely high risk of this pandemic. 266,073 confirmed cases and 11,184 deaths were reported at the global level at that time. India was struggling hard from COVID-19 with a total of 360 confirmed cases and deaths [4]. In India, there is an immediate requirement of hospitals all around the country so temporary hospitals are constructed to overcome loads of COVID-19 patients. Moreover, there are a huge demands of around 1 million ventilators at the peak time of epidemic in India but

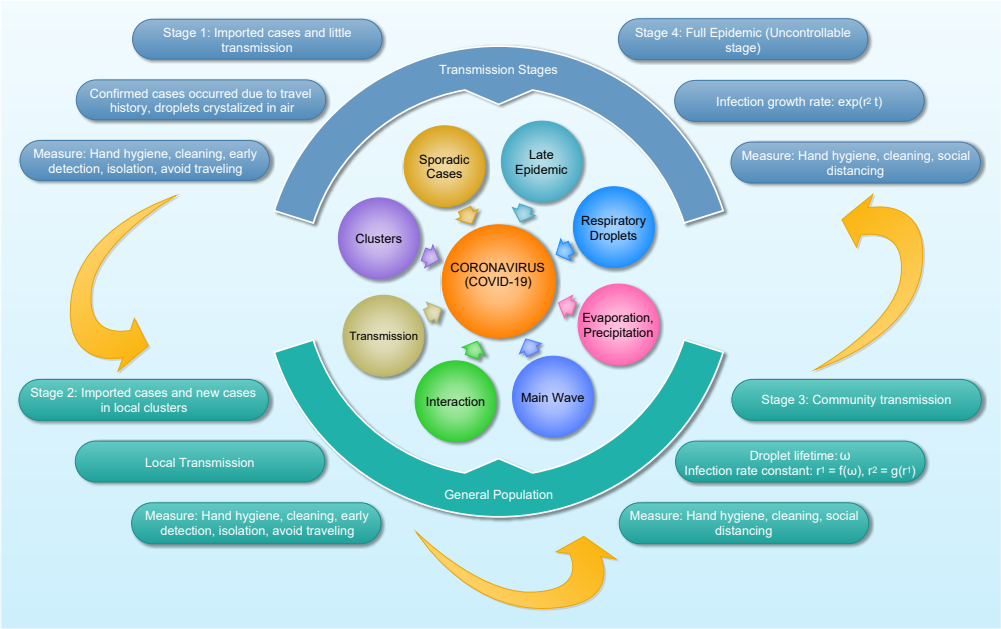

Figure 2: Transmission stages of COVID-19 infection.

Table 1
Summary of the existing state-of-the art schemes.

Schemes	①	②	③	④	⑤	⑥
[5]	Yes	Yes	No	No	No	infection control strategies
[33]	No	No	Yes	Yes	No	real-time live forecasting
[35]	Yes	Yes	Yes	No	Yes	lack of unified databases
[36]	No	Yes	No	Yes	Yes	lack of standard datasets
[37]	No	Yes	Yes	No	No	regulation consideration
[38]	No	Yes	Yes	No	No	security & datasets
[39]	Yes	Yes	No	Yes	No	infection control vaccine
[40]	Yes	No	Yes	Yes	No	standard datasets & suitable prediction model
Our Survey	Yes	Yes	Yes	Yes	Yes	lack of datasets & infection remedies

①: Clinical Characteristics, ②: Detection & Diagnosis, ③: Prediction Models, ④ Datasets, ⑤: Research Gaps, Open Issues & Future Directions, ⑥: Research Challenges.

at present, there are approximate 30K-50K ventilators, similarly it is estimated that the United States has 160K ventilator which is very short as per the patient demands.

1.3. Comparison with the existing surveys

The comparison of the existing state-of-the-art schemes on the COVID-19 is discussed in Table 1.

1.4. Motivation & Scope of the Survey

The motive of this study is not only limited up to the estimation of COVID-19 data but also to provide a solution for the elimination of the coronavirus by applying different prediction models. For remote healthcare management of COVID-19, Machine Learning models are mainly focused in this paper with research challenges of COVID-19 pandemic and their key solutions. As the cases were reporting from all continents of the world and the viral was going on spreading desperately, it becomes a threatening alarm for the whole of mankind. The news was coming from different continents with the adverse effects of this viral disease. The whole world was facing this critical situation and the economy of the world came down due to instant lockdown which ultimately leads to a struggle full life for mankind at the global level. It

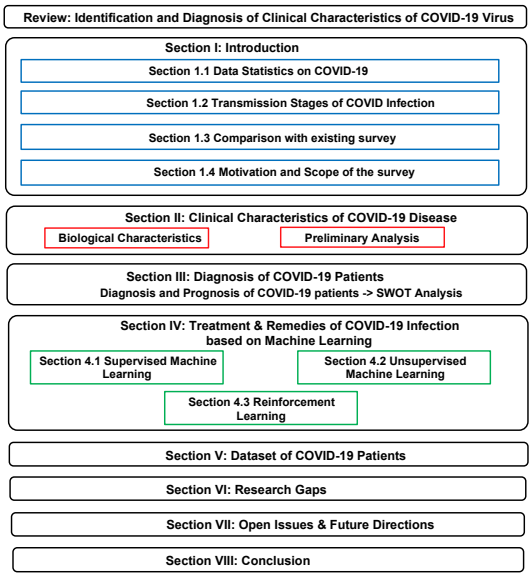

Figure 3: Taxonomy of the literature review on COVID-19.

structured into mind to collect COVID-19 data and to find out solutions using technical applications that may prove helpful in eradication COVID-19. Data collected with the help of WHO reports and worldometer. Motivated from the proposals of the existing work done by the researchers on COVID-19, we presented a literature review on early detection of the symptoms of COVID virus and discuss the performance of the various prediction models based on the machine learning models.

1.5. Paper Structure

Figure 3 presents the taxonomy of the literature review on COVID-19. Section 1 includes the definition and history of COVID-19 followed by the statistics and transmission stages. Then, the comparison with the existing survey, motivation, and scope of the review paper are discussed. Section 2 includes the clinical characteristics of COVID-19 disease. Section 3 considers the diagnosis of COVID-19 patients followed by Section 4 that presents the treatment and remedies adopted for COVID-19 based on machine learning. Section 5 discusses the datasets related to COVID-19 patients. Section 6 focused on the research gaps and Section 7 analyzed the open issues and future directions. Finally, Section 8 concludes the paper.

2. Clinical Characteristics of COVID-19 Disease

Coronaviruses are enfold viruses with a assurance, single-wrecked RNA genome. Coronaviruses possess the largest genomes for RNA viruses having a length ranging between 26 to 32 base pairs (kb). For replication, spike formation, and nucleocapsid of viral, all types of coronaviruses contain specific genes in ORF next regions that conceal proteins. For the attachment and entry of host cells, some glycoprotein spikes are present on the outer surface of coronavirus. The entry receptor for SARS-COV-2 spike glycoprotein is a Human Angiotensin-Converting Enzyme-2 (ACE-2) which is expressed in epithelial cells of the heart, small intestines, and kidneys [5]. In humans and animals, coronaviruses cause intestinal and respiratory infections. While studying on different cases of COVID-19 patients, different infectious categories were reported having decreased volume of the virus in the nasopharyngeal tract and cause organ failure, decreased rate of the virus in an upper and lower tract of the respiratory tract with detection of the virus in plasma, nasopharyngeal shedding of SARAS-COVID-2 in the patients having normal symptoms by early diagnosis.

Table 2 presents the clinical and biological characteristics of the COVID-19 patients. We consider the COVID-19 three patients data such as patients X, Y, and Z, where all these three patients belong to China having a sickness period of 15 days, 26 days, and 23 days, respectively. First patient X was a female having age 30, while the second patient Y is age 48 and the third patient Z is age 80 and both the Y and Z were male. To diagnose COVID-19 patients, the following parameters were taken into account: Age, Chronic Sickness, Province, Travel History, Sickness Period, Diagnosed date, COVID Symptoms, Chest X-ray, Haemoglobin, Platelet Count, Sodium, WBC Count, Urea, Creatinine, Nature Sickness Type, Test Conducted. Symptoms of Patient X and patient Y were detected which include: Fever, Conjunctivitis, Influenza, Cough (mild symptoms) and after the final treatment, both the patients were fully recovered and discharged. Third patient Z having fever, cough, diarrhea, shortness of breath symptoms (severe symptoms), was not able to recover and died on Feb 14, 2020, after a time period of 16 days of hospitalization [5].

3. Diagnosis of COVID-19 Patients

In this section, we diagnose the symptoms of COVID-19 patients. The patient suffering from the COVID-19 virus has the following symptoms such as– fever, cough, difficulty in breathing, repeated shaking with chills, fatigue, muscle pain, headache, sore throat, taste congestion or runny nose nausea, vomiting, and diarrhea. The COVID

Table 2
Clinical and biological characteristics of COVID-19 patients [5].

Clinical Characteristics	COVID-19 Patient X	COVID-19 Patient Y	COVID-19 Patient Z
Patient Sex	Female	Male	Male
Age	30	48	80
Chronic Sickness	Gout Episode	High BP	Thyroid Cancer
Province	Wuhan, China	Shanghai, China	Hubei, China
Travel History	Wuhan to Paris	Shanghai to France	Hubei, China to Europe
Sickness Period	15 days	26 days	23 days
Diagnosed Date	Jan 24, 2020 at Bernard University Hospital	Jan 24, 2020 at Pellegrin University Hospital	Jan 28, 2020
COVID Symptoms	Fever, Conjunctivitis, Influenza	Fever, Cough	Fever, Cough, Diarrhoea, Shortness of breath
Chest X-ray	Bilateral Pneumonia	-	Bilateral Pneumonia
Haemoglobin	13·0	16·9	12·3
Platelet Count	195	182	134
Sodium	142	139	136
WBC Count	3.3	4	8
Urea	2.9	4.4	8
Creatinine	38	68	92
Nature	Nasopharyngeal Swab	Nasopharyngeal Swab	Bronchoalveolar Lavage
Sickness Type	Mild	Mild	Severe
Test Conducted	Hepatitis B and C, Cytomegalovirus, Epstein-Barr virus, Herpes Simplex virus	CT Scan, Ground-glass Opacities, Reticulo-nodular Syndrome, Scarce Alveolar Opacities	Chest x-ray, bilateral alveolar opacities, Renal Replacement Therapy, Acute Respiratory Distress Syndrome, Acute Kidney Injury, Liver Failure, and Sepsis-like Shock
Treatment Outcome	Fully recovered & discharged	Fully recovered, discharged	Not recovered, died on Feb 14, 2020

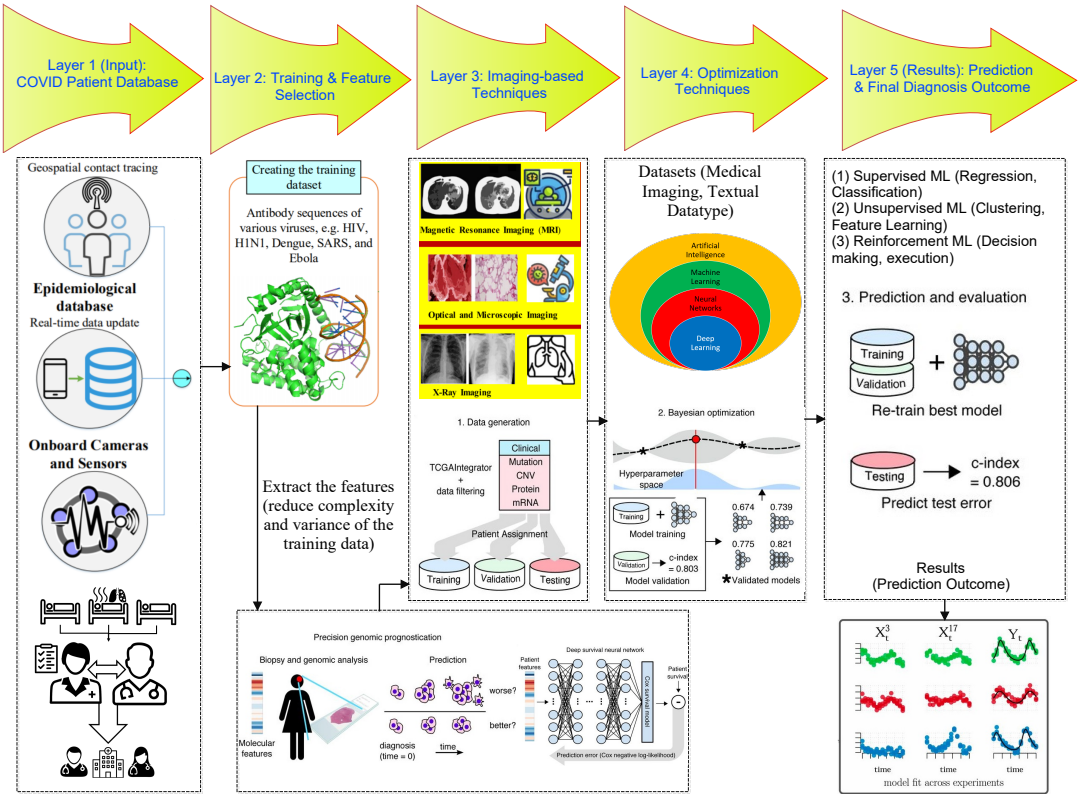

Figure 4: System model of detection & prevention control of COVID-19 virus based on machine learning techniques.

Table 3
Diagnosis and prognosis of COVID-19 patients.

Authors	Disease	Type of Validation	Prognosticators Model	Performance
General Population				
Decaprio *et al.* [6]	COVID-19 pneumonia	training test divided	age, sex, diagnostic history	high
Feng *et al.* [7]	fever	chronological validation	age, temperature, blood pressure, CBC, heart rate	high
Rincon *et al.* [8]	mark population uncertain	10 fold cross validation	fever, shortness of breath	low
Martin *et al.* [9]	patients with suspected COVID-19	outside validation	not specified	high
Zhou *et al.* [10]	paediatric patients with confirm COVID	training test divided	age, sex, WBC count, diabetes, serum albumin	high
Diagnostic Imaging				
Chen *et al.* [11]	suspected COVID-19 pneumonia	training test divided	not appropriate	low
Jin *et al.* [12]	suspected COVID-19 pneumonia	training test divided	not appropriate	high
Shi *et al.* [13]	mark population uncertain	cross validation	location characteristics from imaging	low
Gozes *et al.* [14]	suspected COVID-19 pneumonia	outside validation	not appropriate	high
Wang *et al.* [15]	mark population uncertain	internal and images from other peoples	not appropriate	low
Prognosis				
Zeng *et al.* [16]	confirmed COVID-19 cases with severe disease	unclear fold cross validation	CT feature and laboratory symbols	low
Yuan *et al.* [17]	confirmed COVID-19 cases with mortality	outside validation of current model	clinical storage of CT images	low
Bai *et al.* [18]	patient with mild confirmed COVID-19	uncertain	sign and symptoms, CT images results	high
Xie *et al.* [19]	patients with confirmed COVID-19, death in hospital	outside validation from chinese centre	age, lymphocyte count, LDH	low
Gong *et al.* [20]	patient with severe symptoms confirmed	outside validation from local centres	age, blood urea nitrogen, red blood cell distribution	high
Sarkar *et al.* [21]	patient with COVID-19 symptoms-death vs recovery	outward performance	age, sex, contact with COVID-19 cases	high
Haung *et al.* [22]	patients with severe symptoms, confirmed COVID-19	outward performance	fast respiratory rate, CRP level, LDH level	high
Qi *et al.* [23]	Patients with confirmed COVID-19	cross validation with 5 folds	CT images logistic regression and random forest	low
Lu *et al.* [24]	suspected COVID-19 cases- death within 12 days	not reported	age, CRP	low

Table 4
Strength, Weakness, Opportunities, Threats (SWOT) analysis in diagnosing the symptoms of COVID-19.

SWOT Analysis	Prediction Analysis	Discovery Analysis	Telemedicine Analysis	Containment Analysis	Infodemic Analysis	Economy Disruption Analysis	Social Control Analysis
Strength (S)	time saving	expense effective time period efficient	quick & cost-effective approach for self-diagnose	cost effective approaches	reduces reports	24*7 responses availability	provide hygiene environment that reduces chances of infection
Weakness (W)	generate biasness with arriving data	generate hypothesis discovery	required excellence resources that may not be presented in remote area	required resources not accessible worldwide	NLP is require to categorise false information	require expert supervision	change in behaviour is difficult for elderly population
Opportunities (O)	give a chance for quick planning to diminish loss upon present	no harmful effect on living society	provides opportunity to experts, researchers	develop hygiene and healthy habits	improved learning algorithm can be executed	provides alternate solution to economy	computerized observation with less manpower
Threats (T)	possibility of misleading- adverse prediction	model of treatment and discovery may be differed from real world	raise anxiety in society	instruments Sanitization is necessary after every use	incorrect rejection is possible	rise unemployment percentage	addiction on technology, confidentiality and security issues

patient has to follow the incubation time ranges between 1-15 days after the detection of the virus. Many studies identify the prediction models for valediction of COVID-19 in the broad population. Age, sex, previous hospital history incorporates the predictor factors. Many prognosticator models are applied to different COVID-19 cases to diagnose those. Computed Tomography (CT) images, x-rays, RT-PCR, and many others are used to diagnose the patients in most of the studies. Different prognostic models are used for the patients recognized with COVID-19 by using the many factors which include sign and symptoms, CT features, laboratory symbols, LDH level, and so many others. A more detailed view of diagnostic and prognostic models is specified in Table 3 where all

Table 5
Treatment and remedies of COVID-19 disease.

Innovation	Features	Accuracy	Platform used for Analysis	Expenditure
Testing from waste water	early warning system for COVID-19 cases, highly effective	95%	absorbed the fragments within the sample then extracting the RNA	low
Molecular point-of-care testing	+ve results within 5 minutes and -ve results in 13 minutes, small size, time saving, testing capability is 50,000/day	90%	detects the presence of virus	moderate
Automatic mask creator	fast processing, no infection chances (hygienic)	96%	photoelectric tracking colour code monitoring	40% cheaper
Ruhdaar	well-organized and convenient ventilator	98%	microprocessor controlled mechanical ventilation	low
Jeeva Setu ventilators	small size (Oven size), movable ventilator	90%	delivers 500-550 ml of air per breadth	low
Oura Ring	predict the symptoms of COVID-19 in early 3 days, access through photoplethysmography	90%	artificial intelligence based	moderate
PPEs	fabric appropriate for humid and hot surroundings	95%	high breath ability fabric	low
Microwave Sanitizer Atulya	convenient, fast, insubstantial, operative and efficient	96%	disintegrate viruses and sterilize any surface upto 5 meter depth	moderate
UV Sanitizer device	convenient, ultraviolet ray contactless sanitizer	98%	UV and chemical based sterilization	50% cheaper
Safe Swab	phone booth coronavirus testing, fast and non-violent	92%	health friendly technology	low

the elaborated features and tests used make it clear.

This segment gives an outline of technologies which are recently in use. The topmost attributes of the technologies and their consequences to face the pandemic and to merge the use of these technologies to audit the spreading virus are highlighted in this section. The concerned authorities are informed to poise these technologies which are helpful in eliminating the disease. The work done is described under eight modules such as– prediction analysis, detection analysis, discovery analysis, telemedicine analysis, containment analysis, infodemic analysis, economy disruption analysis, and social control analysis. The strength, weakness, opportunities, and threat (SWOT) analysis in diagnosing the symptoms of COVID-19 is summarized in Table 4.

4. Treatment and Remedies of COVID-19 Infection based on Machine Learning

Figure 4 presents a system model of a machine learning-based scheme for COVID-19 patients. The layered architecture covers 5 layers to diagnose clinical characteristics of COVID-19 patients. The first input Layer is designed for the COVID patients database that can deliver vast packets of data to the main server. The second layer is the training & feature selection, it picks best-suited imaging techniques according to the previous experience of data. The third layer is the imaging-based techniques. CT scan, MRI, PET, X-Ray, Optical, and Digital Microscopic Imaging Techniques are suggested to detect disease by imaging tests. The fourth layer is Optimization Techniques in which Artificial Intelligence (AI), Machine Learning (ML), Convolution Neural Networks (CNN), Deep learning (DL) techniques are used to optimize imaging and textual dataset. The last layer is the output Layer which evaluates and predicts the final diagnosis using the best prediction model.

People from all fields i.e. engineers, doctors, researchers, students, and all others are working on the eradication of COVID-19. Numerous ideas are presented and many innovations are done by them which contains testing from wastewater, molecular point of care testing, automatic mask creator machine, jeeva setu ventilators, qura ring, PPEs, microwave sanitizer atulya, UV sanitizer device and safe swab. Table 5 presents the treatment and solutions that covers the following attributes such as– features, accuracy, platform used for analysis and expenditure of these innovations. For future forecasting of positive confirmed cases, death cases and recovery cases, four prediction models – LR (Linear Regression), LASSO (Least Absolute Shrinkage & Selection Operator), SVM (Support Vector Machine), ES (Exponential Smoothing) are used. Table 6 presents the different prediction models of COVID-19 and their applications. Table 6 covers different applications used by different researchers using various medical standards and approaches. The succeeding applications are used for early detection and diagnosis using radiology images, tracking the outbreak, prediction outcome of COVID-19 patients, protein structure predictions, biomedicine perceptive, pharmacotherapy and drug discovery, and social awareness and control. Chest X-ray images, CT images, the temperature of infected people, infected lung cells, classification of respiratory patterns are the different medical standards used by the researchers using various machine learning models. The performance of these machine learning models are summarized in Table 7. The summarized outcome of the final diagnosis is that the exponential smoothing (ES) machine learning approach performs the best outcome among all

the prediction models [33].

The machine learning model is a mathematical representation of the real-world data the output has been familiar with the certain type of trained designs. In order to perform improved outcomes and performance, developers re-train the existing model. Machine Learning techniques are further divided into three categories i.e., (i) supervised machine learning, (ii) unsupervised machine learning, (iii) reinforcement learning, and (iv) predictive modeling techniques.

4.1. Supervised Machine Learning

The model will be trained with loaded labeled input data. In the labelled dataset, solutions have been given for each dataset. The machine learns from it and makes available solutions for each problem. It predicts the data according to the trained dataset. Supervised Machine Learning is used Classification and Regression methods.

4.1.1. Classification Methods

An extensive range of textual classification methods have been used to estimate social media sentiment and based on their likenesses; these classifiers are assembled into various categories. In this study, two classifiers are mainly focused named as Naïve Bayes Classifier (NBC) and k-Nearest Neighbors (k-NN) [34].

- **Naïve Bayes Classifier (NBC)**: With advanced accuracy to classify text, products, and documents, Naïve Bayes Classifier is a modest and operative method based on Bayes Theorem. To demonstrate this technique, binary input values are used. Bernoulli Naïve Bayes and Multinomial Naïve Bayes are the two models used in Naïve Bayes Classifier which characterizes the frequency and binary features, respectively. To estimate the required parameters, NBCs can be used with restricted size training data. This classifier is operative with real-world data and can deal with dimensionality. Naïve Bayes Classifier probability rule is computed by using the equation.

$$P(A|B) = P(B|A)P(A)/P(B), \quad (1)$$

- **k-nearest Neighbors (k-NN)**: k-Nearest Neighbors, a common non-parametric classifier is easy to implement and appropriate for multi-class complications. Based on the similarity dimension, the k-NN method classifies text and documents very excellently. With slight data, K-NN is a totally efficient and useful algorithm. This classifier works on parallel measurement standards where the approximation of distance and proximity are the aspects used to measure the similarity between two data points. K-NN computes classification based on the bulk of nearest neighbors and their number is resolute within a static radius of an individual point.

4.1.2. Regression Methods

One basic type of Machine Learning is the regression model, it consists of two variables one predictor variable and another is the dependent variable. The regression model analyse the relationship between dependent and independent variables. It predicts the best fit data. Some regression models are listed below.

- **Linear Regression (LR)**: Linear Regression is a valuable and constant classifier, primarily used to predict relationships between continuous variables and used to organize text and documents. The least-squares algorithm method is used to minimize squared alteration between the predicted results and exact classes. To accomplish a regression task, linear regression is used which is a machine learning algorithm based on supervised learning. For future forecasting of positive confirmed cases, death cases, and recovery cases, four prediction models – LR (Linear Regression), LASSO (Least Absolute Shrinkage & amp; Selection Operator), SVM (Support Vector Machine), ES (Exponential Smoothing) are used. The performance of these models is given in Table 7. ES performs the best among all the prediction models.

- The model necessity alterations if the structure is suitable due to non-continuous quantitative variables of predictors or replies. Here, the mathematical model is A, and B are two substitution variables that signify features of documents and class to be evaluated, respectively.

$$A = \alpha_0 + \alpha_1 z_1 + \alpha_2 z_2 + \alpha_3 z_3 + ... + \alpha_n z_n. \quad (2)$$

Estimated value will find $A^{\wedge}$ for a mathematical value $\alpha_0^{\wedge}, \alpha_0^{\wedge}, ...\alpha_n^{\wedge}$ for $\alpha_0, ...\alpha_n$.

- **Support Vector Machine (SVM)**: Support Vector Machine model is used for the regression and classification model of Machine Learning. Due to its increased performance in real-life applications, SVM assists as a foremost tool for data regression and classification. It is a supervised Machine Learning algorithm. SVM can perform proficiently on non-linear classification while executing linear classification. It performs excellently when dimensional spaces are greater than the number of samples.

Table 6
Prediction models of COVID-19 patients and their applications.

Authors	Problem Description	Medical Test	Approaches Used	Applications
Narin *et al.* [25]	abstract radiological structures for appropriate and correct COVID-19 diagnosis, AI is used	chest X-ray images, CT images	CNN models are designed for COVID-19 early detection	detection and diagnosis using radiology images
Wang *et al.* [26]	irregular respiratory pattern categorize that the people infected with COVID-19	classify Respiratory patterns	SEIR model to forecast and time-dependent SIR model is used to track infected peoples	tracking the outbreak
Yan *et al.* [27]	fastidiously and rapidly recognize the risk of death with the help of supervised machine learning	simple and spontaneous clinical test	machine learning based models used to predict accuracy of COVID-19 patients	prediction outcome of COVID-19 patients
Jumper *et al.* [28]	outstanding learning framework is used for image recognition	proteins from genetic order	deep neural network predict characteristics of proteins	protein Structure predictions
Richardson *et al.* [29]	diminish the capability of virus to infested lung cells	infected lung cells	AI models used to examination baricitinib	biomedicine perceptive
Hu *et al.* [30], Ge *et al.* [31]	integrated AI based methods are used for drug discovery	data visualization, novel drug compounds	unsupervised deep learning model is used	pharmacotherapy and drug discovery
Nemati *et al.* [32]	smartphone thermometer kit for retrieving temperature of infected peoples	temperature of infected peoples	AI models is used	social awareness and control by Hi-tech

4.1.3. Clustering Methods

Clustering possess acquires the complementary features as of classification excluding the base. In such a method, we don't have any knowledge about the clusters or segments which we have to spot in our data and what we have to explore. Some effects like clusters, structures and compositions can be unexpectedly emerge in our dataset while using a clustering algorithm. Clustering is used for data compression and generalization of data.

4.2. Unsupervised Machine Learning

Unsupervised Machine Learning finds an unknown pattern in a dataset without pre-existing markers, it is a self-organized technique. There is no need to train output variables only the input dataset is given previously. Unsupervised Machine Learning used Clustering and Reduction Methods for prediction.

4.3. Reinforcement Learning

Reinforcement Learning is an environment interaction-based method. It works on two stages: Start and End stage, there may be different ways to reach an end state. Some examples are driverless cars, robot navigation, self-navigating vacuum cleaner, etc.

4.4. Predictive Modelling- SEIR Model

SEIR stands for Susceptible, Exposed, Infectious, and Removed model. Susceptible are individuals if it exposed infection it becomes hosts, Exposed are individuals that are infected with asymptomatic, Infectious is already infected and it can transmit infection, Removed are individuals who are no lengthier infected and recovered.

5. Datasets of COVID-19 Patients

This section mainly covers the dataset of COVID-19 patients. Table 8 summarized the list of datasets related to COVID-19 patients. The datasets are categorized into two data type i.e., (i) medical imaging and, (ii) textual. The medical imaging data type is further categorized on the basis of CT scans and chest x-ray, while the textual data type is categorized on the basis of COVID-19 case reports, social media data, and scholarly articles. In addition, the features and applications alongwith the trained model used in datasets are elaborated. Different technologies like AI, ML, NLP are used for practical purposes.

6. Research Gaps

After analyzing the existing work by the researchers on COVID-19, the following research gaps have been identified.

- **Less focus on early detection and diagnosis of COVID-19 patients**: High number of patients did not detect at early stages and hence not diagnosed timely. With the rapidly increasing new confirmed cases of COVID day by day, it is necessary to timely diagnose the suspected cases to control the crises of COVID-19.
- **Insufficient utilization of resources like physicians, medical equipment, hospitals**: There is no proper use of resources related to the treatment of COVID-19 patients. It includes many reasons like improperly or less guided doctors who are not able to treat the patients, less amount or sometimes no availability of medical equipment, lack of hospitals up to long distances, and less arrangement of ventilators and proper space in hospitals.

Table 7
Comparison of the performance analysis of various prediction models.

PM	R^2 Score			R^2 Adjusted			MSE			MAE			RMSE		
	C	D	R	C	D	R	C	D	R	C	D	R	C	D	R
LR	0.82	0.95	0.38	0.78	0.94	0.2	1473217659	840310	479124599	29868	725	16977	37872	890	22198
LASSO	0.97	0.86	0.28	0.96	0.8	0.08	234522460.5	3243999.2	1471692577	12105	1429	29875	15416	1799	37395
SVM	0.6	0.52	0.23	0.46	0.4	0.03	5758658971	15909124	1309879561	59983	3241	107877	76017	3901	115603
ES	0.97	0.97	0.98	0.96	0.96	0.98	27960249	658796	6010234	8924	413	1869	17055	820	2373

PM: Prediction Models, LR: Linear Regression, LASSO: Least Absolute Shrinkage and Selection Operator, SVM: Support Vector Machines, ES: Exponential Smoothing, C: Confirmed Cases, D: Death Cases, R: Recovered Cases.

- **Less focus on funding towards drugs and vaccines**: Many researchers and physicians explore that drugs will be a quicker way to tackle the coronavirus pandemic. By drugs and vaccines, protein expansion speeds up in our body for efficient treatment of coronavirus. So the government should be funded towards drugs and vaccination to deal with this deadly disease.
- **No focus on practical and cheap diagnostic**: In this pandemic, proficiency of treatment should be improved and reduce the cost of diagnostic tests. The tests should be available in an affordable range so that every citizen can make himself checked and contribute to eliminating the pandemic permanently by staying healthy and following the instructions given by the government.
- **Insufficient sample size**: Quality of sample size for further treatment and new research (for better solutions) is not adequate. For better treatment and to stop the crisis, it should be considered as a necessary step for a valuable solution.
- **Less focus on ventilators available for COVID patients suffering from breathing**: There is no sufficient availability of ventilators in the hospitals till now. There is a need for the proper arrangement of ventilators in the high range for COVID patients by keeping social distancing. With a proper supply of oxygen, hygienic conditions should be maintained.
- **Less focus on the remote treatment of COVID-19 patients at the time of limited healthcare resources**: The COVID-19 patients have to face a lot of problems due to scarcity of doctors, nursing staff, and other clinical resources. To overcome such issues, some remote controlling techniques should be used which include robots, automatic machines, ventilators accessed with automatic solutions of problems faced by patients such as the supply of oxygen when it gets lowering down, and so on. Such remote-controlled resources can satisfy the needs of patients in the absence of doctors and nursing staff.

7. Open Issues and Future Directions

The open issues, future Directions, and impact of COVID-19 on social life is summarized in Table 9. While presenting different future directions, diverse open issues are challenged which are- 1) In the Transportation sector where vehicles did not move from one place to another due to lockdown. Services were not available due to non-continuous production. 2) In Medical Sector where Emergency health centers and mask creator machine setups and sanitizer machine setups were used. The rest of the issues are described in the table.

The future directions include touchless hi-technology using various approaches, a tracking system for sample collection, the progress of medical services through AI, etc. Social life is affected very badly by a coronavirus. People are facing a lowering of economy, lack of hospital facilities, climate variation, dense population, etc.

8. Conclusion

In this paper, we focused on the identification and diagnosis of the clinical characteristics of COVID-19. Initially, we discuss the data statistics related to COVID-19 alongwith their clinical characteristics. Then, we diagnosis the symptoms and prognosis models of COVID-19 patients. Based on the finding of the research gaps and challenges of COVID-19, we presented the remedies with the help of prediction models such as machine learning, and prediction models. In addition, the datasets model of COVID-19 patients is discussed. Finally, the social impact of pandemic disease, open issues, and future challenges are analyzed. In the future, we implement the forecasting model for the diagnosis of COVID-19 and suggest the methodology for the treatment of the COVID patient.

References

[1] "WHO Coronavirus Disease (COVID-19) Dashboard." (Online Available: https://covid19.who.int/).
[2] H. Swapnarekha, Himansu Sekhar Behera, Janmenjoy Nayak Bighnaraj Naik "Role of Intelligent Com-

Table 8
COVID-19 datasets related to medical imaging & textual datatype.

Datasets Data Type	Features & Applications	Training Approach	Link Available
Medical Imaging Dataset Data Type			
CT Scans	125 COVID-19 images related to MERS-CoV, SARS-CoV, and ARDS , 16 attributes like ID of patient, age of patient, date of admission, and location of patient. Applications: COVID diagnosis.	deep and transfer learning	https://github.com/ieee8023/
CT Scans	275 CT scans extracted from 760 medRxiv and bioRxiv preprints. Application: Diagnosis of COVID	deep CNN with accuracy of around 85%	https://github.com/UCSD-AI4H/COVID-CT
Chest X Ray	Suspect: viral and bacterial pneumonias (MERS, SARS, and ARDS.) Output: binary either +ve or -ve case of Corona virus. Specificity, accuracy and sensitivity achieved by COVID-CAPS are respectively 95.8%, 95.7%, 90% . Application: Diagnosis of COVID.	CNN and SVM, COVID-CAPS model	https://github.com/ieee8023/covid-chestxray-dataset
Chest X Ray	COVID Radiography Database: 219 +ve images of COVID-19, 1341 -ve images of patients and 1345 images of pneuomonia.	CNN model	https://www.kaggle.com/tawsifurrahman/covid19-radiography-database
Textual Dataset Data Type			
COVID-19 Case Report	Diagnose human-to-human transmissions, 3 time-series datasets: daily count of new internationally exported cases, new cases, and the count of infected passengers on evacuation flights. Application:estimating new cases	stochastic transmission	https://github.com/adamkucharski/2020-ncov/
COVID-19 Case Report	utilizes a South Korean benchmark which, announced death rates in sequence of population demographics in order to correct the reported cases. Application: correcting under-report.	statistical	https://www.kaggle.com/lachmann12/correcting-under-reported-covid-19-case-numbers
Social Media Data	UK: (Lock-down, PM in ICU) to report their emotions, dataset with 5000 texts (including 2500 short and 2500 long texts), 2500 was participants. max 240 characters in short tweet-sized text and min 500 characters in long open-ended text. Applications: measuring emotions.	statistical analysis (correlation and regression)	https://github.com/ben-aaron188/covid19worry
Social Media Data	Tweets about pandemic to examine the approaches and awareness, improve public response towards the pandemic and how the response of public changed with time and countries, to evaluate the correct and incorrect information of COVID-19. Applications: Perception.	NLP, data mining	*https://github.com/lopezbec/COVID19_Tweets_Dataset*
Scholarly Articles	initially 28K articles were proposed which have now reached to 52k articles and 41k full texts, various repositories like PMC, BioRxiv and WHO were examine with questioning. Applications: collecting COVID-19 articles.	data extraction, retrieval mining	https://www.semanticscholar.org/cord19/download
Scholarly Articles	In dataset 16K articles on COVID-19 out of 60K articles. Applications: NLP driven search portal.	NLP	https://covidscholar.org

MERS-CoV: Middle East Respiratory Syndrome CoronaVirus, SARS-CoV: Severe Acute Respiratory Syndrome, ARDS: Acute Respiratory Distress Syndrome (ARDS), CNN: Convolutional Neural Network, SVM: Support Vector Machine, COVID-CAPS: Capsule Network Model, NLP: Natural Language Processing.

Table 9
Open issues, future directions, and social impact of COVID-19 on social life.

Impact of COVID-19 on Social Life	Open Issues	Future Directions
Troubled economy, agriculture, organization, and social life	Privacy Laws for operator protection and public privacy	Touchless hi-technology by using AI, IoT, ML, DL, Data Analytics
No availability of hospital facilities, alternate solution followed by peoples like Social distancing, self-quarantine, self-hygiene	Medical database Laws for treatment and proper recommendations	AI advance for smart healthcare to progress medical services
Technology emphasis on value making rather than human administration	Security Laws for proposal and development of technological solutions	Enhance blockchain for a fast response without network delay and with upgraded security
Mostly obstructed measures floating worldwide disparity, climate variation, tense population	Legal Authority and Copyright Laws for distribution content & Personal data	Essential more secure and apparent tracking system for sample collection, drug delivery, and telemedicine.

puting in COVID-19 Prognosis: A State-of-the-Art Review." *Chaos, Solitons & Fractals*, vol. 138, 109947, Sept. 2020.

[3] Punn, Narinder Singh, Sanjay Kumar Sonbhadra, and Sonali Agarwal. "COVID-19 Epidemic Analysis using Machine Learning and Deep Learning Algorithms." medRxiv (2020).

[4] "Coronavirus disease (COVID-19) Situation Report 209." (2020) (Online Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports)

[5] Lescure, Francois-Xavier, Lila Bouadma, Duc Nguyen, Marion Parisey, Paul-Henri Wicky, Sylvie Behillil, Alexandre Gaymard et al. "Clinical and virological data of the first cases of COVID-19 in Europe: a case series." The Lancet Infectious Diseases (2020).

[6] DeCaprio, Dave, Joseph Gartner, Thadeus Burgess, Sarthak Kothari, and Shaayan Sayed. "Building a COVID-19 vulnerability index." arXiv preprint arXiv:2003.07347 (2020).

[7] Feng, Cong, Zhi Huang, Lili Wang, Xin Chen, Yongzhi Zhai, Feng Zhu, Hua Chen et al. "A Novel Triage Tool of Artificial Intelligence Assisted Diagnosis Aid System for Suspected COVID-19 pneumonia In Fever Clinics." (2020).

[8] Lopez-Rincon, Alejandro, Alberto Tonda, Lucero Mendoza-Maldonado, Eric Claassen, Johan Garssen, and Aletta D. Kraneveld. "Accurate identification of sars-cov-2 from viral genome sequences using deep learning." bioRxiv (2020).

[9] Martin, Alistair, Jama Nateqi, Stefanie Gruarin, Nicolas Munsch, Isselmou Abdarahmane, and Bernhard Knapp. "An artificial intelligence-based first-line defence against COVID-19: digitally screening citizens for risks via a chatbot." bioRxiv (2020).

[10] Zhou, Ying, Zhen Yang, Yanan Guo, Shuang Geng, Shan Gao, Shenglan Ye, Yi Hu, and Yafei Wang. "A new predictor of disease severity in patients with COVID-19 in Wuhan, China." medRxiv (2020).

[11] Chen, Jun, Lianlian Wu, Jun Zhang, Liang Zhang, Dexin Gong, Yilin Zhao, Shan Hu et al. "Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: a prospective study." MedRxiv (2020).

[12] Jin, Cheng, Weixiang Chen, Yukun Cao, Zhanwei Xu, Xin Zhang, Lei Deng, Chuansheng Zheng, Jie Zhou, Heshui Shi, and Jianjiang Feng. "Development and Evaluation of an AI System for COVID-19 Diagnosis." medRxiv (2020).

[13] Shi, Feng, Liming Xia, Fei Shan, Dijia Wu, Ying Wei, Huan Yuan, Huiting Jiang, Yaozong Gao, He Sui, and Dinggang Shen. "Large-scale screening of covid-19 from community acquired pneumonia using infection size-aware classification." arXiv preprint arXiv:2003.09860 (2020).

[14] Gozes, Ophir, Maayan Frid-Adar, Hayit Greenspan, Patrick D. Browning, Huangqi Zhang, Wenbin Ji, Adam Bernheim, and Eliot Siegel. "Rapid AI development cycle for the coronavirus (covid-19) pandemic: Initial results for automated detection & patient monitoring using deep learning ct image analysis." arXiv preprint arXiv:2003.05037 (2020).

[15] Wang, Shuai, Bo Kang, Jinlu Ma, Xianjun Zeng, Mingming Xiao, Jia Guo, Mengjiao Cai et al. "A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19)." MedRxiv (2020).

[16] Zeng, Lijiao, Jialu Li, Mingfeng Liao, Rui Hua, Pilai Huang, Mingxia Zhang, Youlong Zhang et al. "Risk assessment of progression to severe conditions for patients with COVID-19 pneumonia: a single-center retrospective study." medRxiv (2020).

[17] Yuan, Mingli, Wen Yin, Zhaowu Tao, Weijun Tan, and Yi Hu. "Association of radiologic findings with mortality of patients infected with 2019 novel coronavirus in Wuhan, China." PloS one 15, no. 3 (2020): e0230548.

[18] Bai, Xiang, Cong Fang, Yu Zhou, Song Bai, Zaiyi Liu, Liming Xia, Qianlan Chen et al. "Predicting COVID-19 malignant progression with AI techniques." (2020).

[19] Jianfeng, Xie, Hungerford Daniel, Chen Hui, T. Abrams Simon, Li Shusheng, Wang Guozheng, Wang Yishan, Kang Hanyujie, Bonnett Laura, and Zheng Ruiqiang. "development and external validation of a prognostic multivariable model on admission for hospitalized patients with covid-19." https://www. medrxiv. org/con-

tent/medrxiv/early/2020/03/30/2020.03. 28.20045997. full. pdf (2020).

[20] Gong, Jiao, Jingyi Ou, Xueping Qiu, Yusheng Jie, Yaqiong Chen, Lianxiong Yuan, Jing Cao et al. "A tool to early predict severe 2019-novel coronavirus pneumonia (COVID-19): a multicenter study using the risk nomogram in Wuhan and Guangdong, China." medRxiv (2020).

[21] Sarkar, Jit, and Partha Chakrabarti. "A Machine Learning Model Reveals Older Age and Delayed Hospitalization as Predictors of Mortality in Patients with COVID-19." medRxiv (2020).

[22] Huang, Huang, Shuijiang Cai, Yueping Li, Youxia Li, Yinqiang Fan, Linghua Li, Chunliang Lei et al. "Prognostic factors for COVID-19 pneumonia progression to severe symptom based on the earlier clinical features: a retrospective analysis." medRxiv (2020).

[23] Qi, Xiaolong, Zicheng Jiang, Qian Yu, Chuxiao Shao, Hongguang Zhang, Hongmei Yue, Baoyi Ma et al. "Machine learning-based CT radiomics model for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection: A multicenter study." medRxiv (2020).

[24] Lu, Jiatao, Shufang Hu, Rong Fan, Zhihong Liu, Xueru Yin, Qiongya Wang, Qingquan Lv et al. "ACP risk grade: a simple mortality index for patients with confirmed or suspected severe acute respiratory syndrome coronavirus 2 disease (COVID-19) during the early stage of outbreak in Wuhan, China." (2020).

[25] Narin, Ali, Ceren Kaya, and Ziynet Pamuk. "Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks." arXiv preprint arXiv:2003.10849 (2020).

[26] Wang, Shuai, Bo Kang, Jinlu Ma, Xianjun Zeng, Mingming Xiao, Jia Guo, Mengjiao Cai et al. "A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19)." MedRxiv (2020).

[27] Yan, Li, Hai-Tao Zhang, Jorge Goncalves, Yang Xiao, Maolin Wang, Yuqi Guo, Chuan Sun et al. "A machine learning-based model for survival prediction in patients with severe COVID-19 infection." MedRxiv (2020).

[28] Senior, Anderw, John Jumper, Demis Hassabis, and Pushmeet Kohli. "AlphaFold: Using AI for scientific discovery." DeepMind. Recuperado de: https://deepmind. com/blog/alphafold (2018).

[29] Richardson, Peter, Ivan Griffin, Catherine Tucker, Dan Smith, Olly Oechsle, Anne Phelan, and Justin Stebbing. "Baricitinib as potential treatment for 2019-nCoV acute respiratory disease." Lancet (London, England) 395, no. 10223 (2020): e30.

[30] Hu, Fan, Jiaxin Jiang, and Peng Yin. "Prediction of potential commercially inhibitors against SARS-CoV-2 by multi-task deep model." arXiv preprint arXiv:2003.00728 (2020).

[31] Ge, Yiyue, Tingzhong Tian, Sulin Huang, Fangping Wan, Jingxin Li, Shuya Li, Hui Yang et al. "A data-driven drug repositioning framework discovered a potential therapeutic agent targeting COVID-19." bioRxiv (2020).

[32] Nemati, Ebrahim, Md Mahbubur Rahman, Viswam Nathan, Korosh Vatanparvar, and Jilong Kuang. "A comprehensive approach for cough type detection." In 2019 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), pp. 15-16. IEEE, 2019.

[33] Rustam, Furqan, Aijaz Ahmad Reshi, Arif Mehmood, Saleem Ullah, Byungwon On, Waqar Aslam, and Gyu Sang Choi. "COVID-19 Future Forecasting Using Supervised Machine Learning Models." IEEE Access (2020).

[34] Samuel, Jim, G. G. Ali, Md Rahman, Ek Esawi, and Yana Samuel. "Covid-19 public sentiment insights and machine learning for tweets classification." Information 11, no. 6 (2020): 314.

[35] Agarwal, Sonali, Narinder Singh Punn, Sanjay Kumar Sonbhadra, P. Nagabhushan, K. K. Pandian, and Praveer Saxena. "Unleashing the power of disruptive and emerging technologies amid COVID 2019: A detailed review." arXiv preprint arXiv:2005.11507 (2020).

[36] Pham, Quoc-Viet, Dinh C. Nguyen, Won-Joo Hwang, and Pubudu N. Pathirana. "Artificial Intelligence (AI) and Big Data for Coronavirus (COVID-19) Pandemic: A Survey on the State-of-the-Arts." (2020).

[37] Samuel, Jim, G. G. Ali, Md Rahman, Ek Esawi, and Yana Samuel. "Covid-19 public sentiment insights and machine learning for tweets classification." Information 11, no. 6 (2020): 314.

[38] Kumar, Aishwarya, Puneet Kumar Gupta, and Ankita Srivastava. "A review of modern technologies for tackling COVID-19 pandemic." Diabetes & Metabolic Syndrome: Clinical Research & Reviews (2020).

[39] Punn, Narinder Singh, Sanjay Kumar Sonbhadra, and Sonali Agarwal. "COVID-19 Epidemic Analysis using Machine Learning and Deep Learning Algorithms." medRxiv (2020).

[40] Hussain, Adedoyin Ahmed, Ouns Bouachir, Fadi Al-Turjman, and Moayad Aloqaily. "AI techniques for covid-19." IEEE Access 8 (2020): 128776-128795.