

# Big Data Analytics in COVID-19

Subjects: **Computer Science, Information Systems**

Contributor: Abdullah Almuhaideb

The COVID-19 epidemic has caused a large number of human losses and havoc in the economic, social, societal, and health systems around the world. Controlling such epidemic requires understanding its characteristics and behavior, which can be identified by collecting and analyzing the related big data. Big data analytics tools play a vital role in building knowledge required in making decisions and precautionary measures. However, due to the vast amount of data available on COVID-19 from various sources, there is a need to review the roles of big data analysis in controlling the spread of COVID-19, presenting the main challenges and directions of COVID-19 data analysis, as well as providing a framework on the related existing applications and studies to facilitate future research on COVID-19 analysis.

artificial intelligence (AI)

big data

big data analytics

2019 novel coronavirus disease (COVID-19)

healthcare

## 1. Introduction

On 30 January 2020, the World Health Organization (WHO) declared the spread of the COVID-19 pandemic as a cause of concern and called for raising the level of health emergencies. Afterward, the government of the Kingdom of Saudi Arabia urgently took several strict measures to limit the spread of the pandemic within the regions of Saudi Arabia <sup>[1][2]</sup>. The Saudi Ministry of Health (MoH) and many other countries have implemented WHO recommendations related to the identification and isolation of suspected COVID-19 cases.

Nevertheless, the pandemic has spread dramatically, with the number of infected people over 82 million, and the number of deaths exceeding one million <sup>[3]</sup>. The rapid spread of the pandemic, with its continuous evolving patterns and the difference in its symptoms, makes it more difficult to control. Moreover, the pandemic has affected health systems and the availability of medical resources in several countries around the world, contributing to the high death rate <sup>[4]</sup>.

A regular monitoring and remote detection system for individuals will assist in the fast-tracking of suspected COVID-19 cases. Moreover, using such systems will generate a huge amount of data, which will provide many opportunities for applying big data analytics tools <sup>[5]</sup> that are likely to improve the level of healthcare services. There are a large number of open-source software such as the big data components for the Apache project <sup>[6]</sup>, which are designed to operate in a cloud computing and distributed environment to assist in the development of big data-based solutions. Furthermore, there are several key characteristics of big data called the Six V's <sup>[7]</sup>, namely,

Value, Volume, Velocity, Variety, Veracity, and Variability. However, the original definition of the big data key characteristics considers only three Vs, namely Volume, Velocity, and Variety [8].

The big data characteristics apply to data acquired from the healthcare sector, which increases the tendency to use big data analysis tools to improve sector services and performance. There are wide applications of big data analytics in the healthcare sector, including genomics [9], drug discovery and clinical research [10], personalized healthcare [11], gynecology [12], nephrology [13], oncology [9][12], and several other applications found in the literature. However, in this paper, we present the contributions of the most important review papers found in the literature that cover the field of big data in healthcare. We also investigate the opportunities and challenges for applying big data analytics tools to COVID-19 data and provide findings and future directions at the end of the paper.

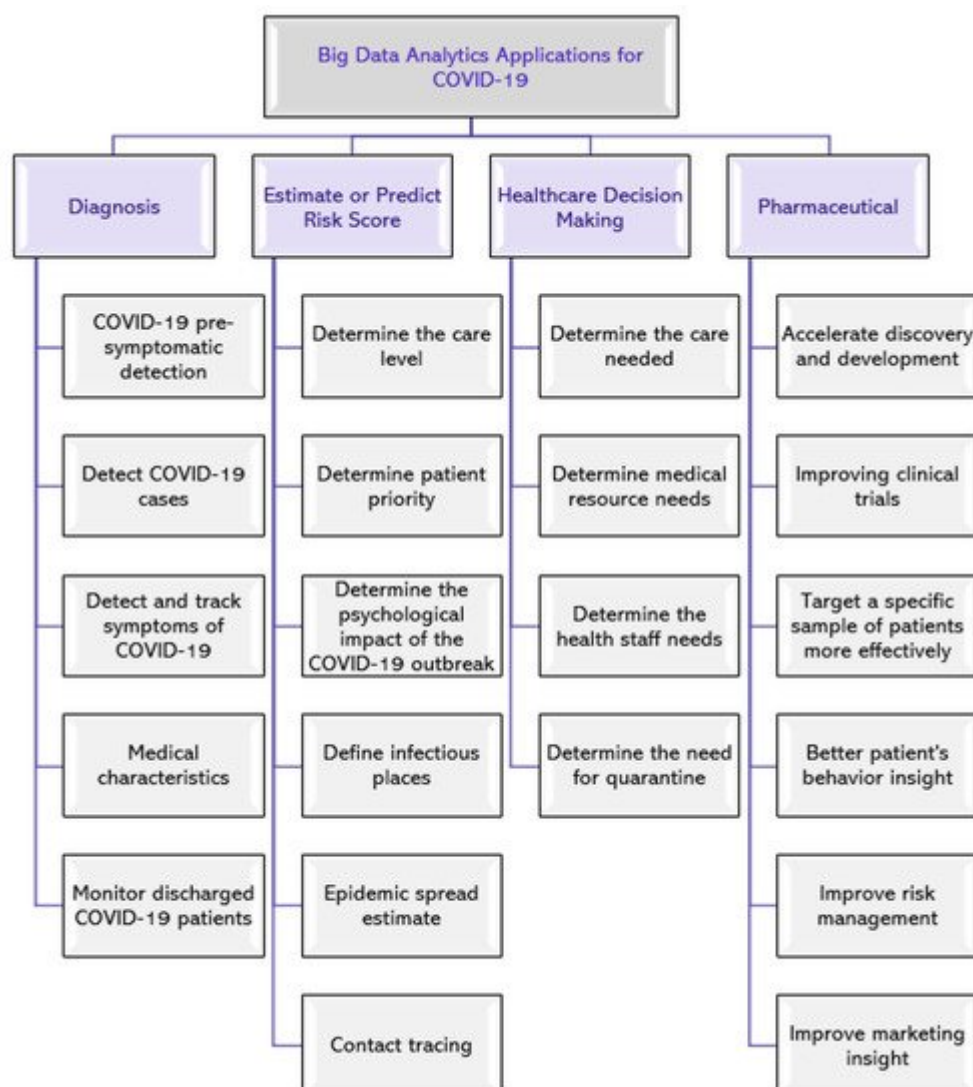
Promising wearable technology is expected to be one of the primary sources of health information, given its widespread availability and acceptance by people. Based on a survey conducted in January 2020, 88% of 4600 subjects included in the study indicated a willingness to use wearable technology to measure and track their vital signs. While 47% of chronically ill patients and 37% of non-chronically ill patients reported a willingness to blindly share their health information with healthcare research organizations. Of the same group, 59% said they would likely use artificial intelligence (AI)-based services to diagnose their health symptoms [14]. People sharing such data routinely will greatly increase the volume of data, which calls for planning to design and implement data analysis tools and models in this sector.

Several studies used big data for sentiment analysis, such as Reference [15], which linked between social media behavior and political views, opinions, and expressions. The study consisted of a representative survey conducted on 62.5% of adults from Chile and it showed the huge effect of social media on changing people's opinions regarding political views and elections. Similarly, the authors of Reference [16] had studied how the management responding to customer satisfaction online review affects the choice of the customers for some facilities or hotels. It showed a positive correlation between the response and customer satisfaction. The authors of Reference [17] had reviewed the classification techniques, including deep and convolutional, to identify the writer from their handwriting. They discussed several challenges in identification related to language characteristics, scripts, and the lack of datasets. Also, the authors of Reference [18] had reviewed and analyzed the latest papers about big data analytics latest developments, capabilities, and profits. Their study showed that big data can support business industries in many functionalities including prediction, planning, managing, decision-making, and traceability. The limitation of their study is the data sources, which were hard to find due to privacy and conservation of the information. Moreover, the authors of Reference [19] had surveyed numerous papers about mathematical models to improve the efficiency in detecting and predicting COVID-19. Their survey suggested using artificial intelligence to detect COVID-19 cases, big data to trace cases, and nature-inspired computing (NIC) to select suitable features to increase the accuracy of detection. Some surveys studied heart-related diseases and suggested some recommendations and guidelines, such as Reference [20], to help people in understanding heart failure causes, symptoms, and the most affected group. They declared that heart failure can escalate the patient's injuries, especially the ones with serious illnesses.

Analyzing health data in real-time with the utilization of AI techniques will have a vital role in predictive and preventive healthcare. For example, it will help predict the sites of infection and the flow of the virus. It will also help in estimating the needs of beds, healthcare specialists, and medical resources during such pandemic crises as well as in the diagnosis and characterization of the virus <sup>[21]</sup>.

## 2. Applications of Data Analytics in COVID-19

The spread of the global pandemic, COVID-19, has generated a huge and varied amount of data, which is increasing rapidly. This data can be used by applying big data analytics techniques in multiple areas, including diagnosis, estimate or predict risk score, healthcare decision-making, and pharmaceutical industry <sup>[22]</sup>. [Figure 1](#) shows examples of potential application areas.



**Figure 1.** Potential application areas of big data analytics for COVID-19.

In the following subsections, we present several examples of COVID-19 data utilization from the literature with a primary focus on reviewing studies that have provided solutions to control the COVID-19 pandemic and fall within

one of the three areas, namely (1) diagnosis, (2) estimate or predict risk score, and (3) healthcare decision-making . We also summarize the data analysis techniques and the data type used for each study in [Table 1](#).

**Table 1.** Data analysis technique, type, source, and findings of the existing studies.

Area	Ref	Aim	Technique	Used Data Type	Data Source	Findings
Diagnosis	[23]	Develop a diagnosis model for COVID-19 detection and diagnosis of symptoms to define appropriate care measures	Best Worst Method (BWM)	Symptoms and CT scans	Body sensors	The model can differentiate COVID-19 from four other viral chest diseases with 98% accuracy
	[24]	Design a medical device to detect and track respiratory symptoms of COVID-19	N/A	Symptoms	Headsets and mobile phone	The approach provided good and stable results and can be expanded to include more sensors to detect other COVID-19 symptoms
	[25]	Develop a remote patient monitoring program (RPM) for discharged COVID-19 cases	The mixed-effects logistic regression model	Demographics, medical data	The remote monitoring program, pulse oximeter, and thermometer	RPM provides scalable remote monitoring capabilities and decreases readmission risk
	[26]	Investigate smartwatches usefulness in pre-symptoms COVID-19 detection	Two anomaly detection models (RHR-Diff and HROS-AD)	Demographics, activity, medical data, COVID-19 status	Smartwatches and MyPHD mobile app	Respiratory infections can be detected through activity tracking and health monitoring via wearable devices
	[27]	Identify symptoms associated with positive COVID-19 cases	Principal component analysis (PCA), and logistic regression model	Demographics, medical data	Screening via phone and COVID-19 PCR test	Fever, anosmia/ageusia, and myalgia were the strongest signs of positive COVID-19 cases,

Area	Ref	Aim	Technique	Used Data Type	Data Source	Findings
						while no symptoms were limited to nasal congestion/sore throat associated with negative cases
	[28]	Determine the clinical characteristics and outcomes of COVID-19 patients in the NY area	N/A	Demographics, medical data, COVID-19 status	Northwell Health system	The common comorbidities were obesity, hypertension, and diabetes. From outpatients or dead patients ( $n = 2634$ ): 21% died, 14.2% were treated in the ICU, 12.2% received MV, and 3.2% were treated with kidney replacement
	[29]	Distinguish COVID-19 cough sound from other respiratory diseases through crowd source data	Logistic Regression (LR), Gradient Boosting Trees, and Support Vector Machines (SVMs)	Demographics, medical data, COVID-19 data	Web app and Android app	Wet and dry cough are the common symptoms of positive COVID-19 cases, whereas chest tightness and the lack of smell are the common combination symptoms
	[30]	Discuss the importance of developing complementary technologies to diagnose and monitor COVID-19 infections	N/A	Activity data, medical data	Sensors	Recommend deploying advanced wearable technologies configured to directly address needs in COVID-19 monitoring and noticing the symptoms

Area	Ref	Aim	Technique	Used Data Type	Data Source	Findings
	[31]	Identify the clinical characteristics of COVID-19 to help in mapping the disease and guiding pandemic management	N/A	Demographics, medical data, COVID-19 status, travel data	Health Electronic Surveillance Network (HESN) database for all Saudi Arabia regions	Fever and cough were common symptoms in the study sample
	[32]	Employing a two-stage cascading platform to enhance the accuracy of machine learning models	Progressive machine learning technique merged with Spark-based linear models, Multilayer Perceptron (MLP), and LSTM	Medical data	Cardiac Arrhythmia Database. Uniform Resource Locator (URL) Reputation Dataset from University of California Irvine Machine Learning (UCI ML) Repository	Using an improved algorithm with two-step data analysis platforms can increase accuracy in lower computation time
	[33]	Analyzing the dense layers among the convolutional network can help to increase the accuracy of classification of images for diabetic retinopathy	Deep learning model	Medical data, Demographics data	The Messidor-2 dataset from the hospital	Using improved programming technology can enhance accuracy
	[34]	Analyze the effects of COVID-19 on patients with cardiovascular disease	Generalized linear mixed model	Demographics, medical data, COVID-19 status	HERs from General Hospital of Central Theatre Command in Wuhan, China	Middle-aged and elderly heart patients are most likely to have COVID-19, whereas new-onset hypertension and heart injury are common

Area	Ref	Aim	Technique	Used Data Type	Data Source	Findings
						complications of severe COVID-19 cases
Estimate or Predict Risk Score						Cardiac function and vital signs should be monitored in COVID-19 patients, especially those with hypotension, pericardial effusion, or severe myocardial injury
	[35]	Specify the effect of COVID-19 on the cardiovascular system	The multi -factor logistic regression model	Demographics and medical data	HERs	
	[36]	Develop and validate a risk score to predict adverse events of suspected COVID-19 patients	Least absolute shrinkage and selection operator (LASSO) and logistic regression models	Demographics and medical data	15 EDs in Southern California	COVAS score can help physicians to identify patients who may experience a serious event within 7 days
	[37]	Discover unregistered suspected COVID-19 patients and infectious places	SIR and $\theta$ -SEIHRD mathematical models	Demographics and COVID-19 data	IoT-based system and GPS	The proposed system helps identify people who had close contact with COVID-19 patients
	[38]	Verify if the COVID-19 virus can be transmitted through indirect contact	N/A	Demographics, medical, environmental, and other data	Guangzhou CDC database and sample collection	The virus can survive for a short period on surfaces, allowing indirect transmission of infection to uninfected people
	[39]	Identify the COVID-19 outbreak impact on the psychological side	Bivariate linear regression	Demographics, medical, social data	Online questionnaire	The COVID-19 outbreak has a significant mental impact on people

Area	Ref	Aim	Technique	Used Data Type	Data Source	Findings
	[40]	Analyze the risk of tuberculosis skin on getting infected by tuberculosis	Statistical	Medical data, Demographics data	Public source	The tuberculin skin can increase the infection by up to 20%
	[41]	Predict the course of the COVID-19 epidemic to design a control strategy	A designed mathematical model called SIDARTHE	Demographics, medical, environmental data	Public data from Italian MoH and Italian Civil Protection	Social distancing measures and lockdowns are necessary and effective, and precautionary measures for COVID-19 can only be relieved when tests are conducted on a large scale and a mechanism for contact tracing is in place
Healthcare Decision-Making	[42]	Evaluate the effectiveness of COVID-19 control measures	C-SEIR model(mathematical model of disease transmission dynamics)	Confirmed COVID-19 data	Public data sources	Quarantine measures have an effective role in containing COVID-19, but they are economically expensive
	[43]	Develop a patient monitoring platform to directly provide the necessary care	N/A	Demographics, medical, COVID-19 data	Online questionnaire via patient monitoring program	Analyzing patient monitoring data helps to know the risk score to determine the care required, allowing optimal consumption of medical resources
	[44]	Provide a platform for data collection and analysis to estimate disease incidence to	Weighted prediction model	Demographics, medical, COVID-19, and other data	Mobile app	Existing data collection methods can be repurposed to track and obtain real-time data for the population



Area	Ref	Aim	Technique	Used Data Type	Data Source	Findings
		develop risk mitigation strategies and resource allocation				during any rapid global health crisis
	[45]	Identify the regional distribution of the spread of infection and the percentage of healthcare consumption in each region	N/A	Demographics, medical, and other data	Mobile app	Can rely on the mobile app to perform self-assessment and data collection that can be displayed on an interactive map and linked to the results of the COVID-19 test results to support decision-makers and healthcare providers in making decisions
	[46]	Forecast the census and ventilators requirements for a specific hospital	Weibull and conditional distributions (analytical model)	Statistical data	COVID-19 hospitalized patient records	The model can predict the census and the required number of MV in one, three, and seven days after the simulation run date
	[47]	Estimate the need for health services and the number of daily deaths over the next 4 months from the date of the study	Statistical model	COVID-19 and other data	WHO websites and local and national authorities in the US states	The model predicts an increased death rate and demand for medical beds, ICU, and MVs
	[48]	Prove that the three clinical variables: age, fever, tachypnea, can be used to predict the	EHRead from Savana [49], and deep learning convolutional neural network classification methods (Prediction model)	Demographics, medical data	EHRs of the hospitals within the Servicio de Salud de Castilla-La Mancha	The most common symptoms of male COVID-19 with an average age of 58.2 years who were

RT-PCR)  
countries

experienced increased demand for diagnosing suspected COVID-19 cases, which exceeded the available local testing capacity. Therefore, several researchers have proposed alternative solutions for the COVID-19 RT-PCR diagnosis test, including the following.

Area	Ref	Aim <sup>[23]</sup>	Technique	Used Data Type	Data Source	Findings
		need to admit COVID-19 patients into the ICU			(SESCAM) Healthcare Network in Castilla-La Mancha, Spain	admitted to ICU are coughing, fever, and shortness of breath, while those between 40 and 79 years of age are likely to be admitted to the ICU if they suffer from rapid breathing
	[19]					
	[24] [50] [25]	Pre-risk assessment of the epidemic in Italy and identification of high-risk areas	a-priori effect of hazard and vulnerability model (a-priori E_H_V)	Statistical and environmental data	Data from Italian Ministry of Economic Policy Planning and Coordination, Italian Ministry of Health website, WHO, Italian Ministry of Agriculture, and ISTAT database	The risk of a pandemic is higher in some northern regions of Italy and the policy model developed can help policymakers make decisions
	[51]	Estimate the remaining period before consuming the operational capacity of the hospital and its resources	Monte Carlo simulation, SIR model, and COVID-19 Hospital Impact Model (CHIME)	Statistical data	Academic health system for three hospitals in the Philadelphia region	The model can help in making proactive decisions

found that activity tracking and health monitoring using wearable devices can help in early detection of respiratory infections.

Since the COVID-19 symptoms have not been fully identified and due to the changing nature of COVID-19, some studies have focused on identifying the medical characteristics and symptoms associated with positive COVID-19 cases. The study in Reference [27] focused on identifying the symptoms associated with the positive results of the COVID-19 examination, and it was conducted on a group of healthcare workers (HCWs). Initial screening was performed by phone, and a COVID-19 PCR test was also performed for each HCW to identify symptoms associated with each case. The study found that the most common symptoms of positive COVID-19 cases were fever, myalgia, and anosmia/ageusia, while the negative cases mostly have no symptoms, or the symptoms are limited to nasal congestion and sore throat.

The study in Reference [28] aimed to determine the clinical characteristics and outcomes of 5700 hospitalized patients with COVID-19 in the NY area. However, the study included non-critically ill patients and the follow-up time was limited.

Another study [29] proposed a website and Android app to separate a COVID-19 cough sound from other respiratory sounds with the aid of crowdsourcing data from about 7000 unique users (more than 200 of whom reported a recent positive test for COVID-19). Their proposed method employed Logistic Regression (LR), Gradient Boosting Trees, and Support Vector Machines (SVMs) classifiers to distinguish the cough sound data based on gender, age, and symptoms. Also, their classifiers distinguish the user based on other features, such as whether they are asthmatic patients, smokers, or healthy. Their app asks the user to cough from three to five times then repeat the process every two days to update the user's health status. Their method proved that a COVID-19 cough can be distinguished from other lung diseases coughs from the sound of the cough combined with breathing sound to screen the disorder. It achieved 82% Area Under the Curve (AUC) in identifying the cases that tested positive for COVID-19. They recommended more studies in the field to specify more characteristics of a COVID-19 cough sound to make it more distinguishable from other respiratory sounds.

The authors in Reference [30] declared the importance of using complementary technologies such as on-body sensors for diagnosing and monitoring COVID-19 infections. They stated that clinical devices are more reliable and provide more functions than smartwatches since these devices are distributed in different areas of the human body to detect different body signals. A thin, soft sensor with a high-bandwidth accelerometer and a precision temperature sensor placed on the neck is very important to record respiratory activity from cough frequency, intensity, and duration to respiratory rate and effort, to high-frequency respiratory features associated with wheezing and sneezing. Also, they recommended machine learning and predictive algorithms to help to diagnose and monitor COVID-19.

In Reference [31], researchers emphasized on the importance of identifying the characteristics of COVID-19 among patients of Saudi Arabia in managing the pandemic. The study included 1519 cases where data related to their ages, genders, vital signs, public data, and clinical examinations were collected. Their test was conducted based on the quantitative RT-PCR approach, which is the protocol established by the World Health Organization. After the data was gathered, it was entered into electronic sheets with distinct data collectors, and data was analyzed with Statistical Package for Social Sciences program, version 24 (SPSS-24). The statistics manifested that the most common symptoms of COVID-19 are cough and fever, with 89.4% and 85% presence in reported positive cases, respectively. Also, it confirmed that the most infected patients' demographics include elder males, severe cardiac condition patients, and diabetic patients.

The authors in Reference [32] had utilized machine learning techniques along with spark-based linear models, Multilayer Perceptron (MLP), and Long Short-Term Memory (LSTM) with a two-stage cascading platform to enhance the prediction accuracy in different datasets. They applied their method on two datasets for cardiac arrhythmia and resource locator, so their model performed with higher accuracy and lower computation time. Thus, the authors in Reference [33] had proposed a computer program method to aid the classification model to analyze

the retinal image of diabetic retinopathy to investigate its effect among adults in causing blindness. It proved that the focused connection among layers of the convolutional network assists the accuracy of the classification result.

The retrospective, observational study in Reference [34] conducted a statistical analysis to show the cardiovascular implications of COVID-19 on the patients. The study was performed on 116 patients who tested positive for COVID-19. The data was clinically collected and tested to extract clinical symptoms and signs, chest computed tomography, treatment measures, and medical records. The statistical analysis was performed on the data to reveal similar results as those reported by Reference [31], where the common symptoms were fever and dry cough, and the elder or middle-aged males, heart injury patients, hypertension patients, and diabetics were the most infected populations.

## 2.2. Estimate or Predict Risk Score

Estimating the risk score helps in determining the care level and priority for each patient with an insight to the necessary proactive measures. In the following section, we present the studies that cover this area.

In Reference [35], the authors aimed to validate a hypothesis that COVID-19 infection could lead to serious cardiovascular diseases or maybe worse. They utilized statistical analysis by employing a multi-factorial logistic regression model to analyze COVID-19-related causes. The study was conducted on 54 patients with different ages, genders, and vital signs, where 39 were diagnosed as severe COVID-19 cases and 15 as critical COVID-19 cases. The data was collected clinically from the patients with attached vital sign measurement devices updated every four hours. Results showed that elder males, diabetic patients, and hypotension patients are more likely to develop a serious heart-related condition and need more care. Their proposed study is limited due to the small sample size, and they suggested a higher sample size to conduct a more appropriate study and verify the results.

The authors in Reference [36] are interested in developing and validating the risk score to predict adverse events among patients suspected of having COVID-19. They conducted a retrospective cohort study of adult visits to the emergency department. The study concluded that the primary outcome was death or no respiratory decompensation within 7 days. To derive the risk score, they used the Least Absolute Shrinkage and Selection (LASSO) and Logistic Regression models. They concluded that the COVID-19 Acuity Score (COVAS) can assist in decision-making to discharge patients during the COVID-19 pandemic. They also reported the derivation and validation metrics of cohorts and subgroups with pneumonia or COVID-19 diagnosis.

The authors in Reference [37] proposed an Internet of Things (IoT) based system to discover unregistered COVID-19 patients, as well as infectious places. This would help the responsible authorities to disinfect contaminated public places and quarantine the infected persons and their contacts even if they did not have any symptoms. The newly confirmed and recovered cases would be recorded in the system by the healthcare staff, while the geolocation data will be collected automatically by Global Positioning System (GPS) technology in the IoT devices. The authors discussed how their proposed system could be utilized to apply three different prediction mathematical models, namely the  $\theta$ -SEIHRD model, Susceptible-Infected-Recovered (SIR) model, and Susceptible-Exposed-Infectious-Removed (SEIR) model.

Another study [38] demonstrated the possibility of transmitting the COVID-19 virus through indirect contact, like touching surfaces contaminated with the droplets of an infected person. Therefore, it was recommended that paying attention to personal hygiene and disinfection of public places could possibly reduce the incidence.

Furthermore, researchers also [39] conducted a cross-sectional study to show the impact of the COVID-19 outbreak on the psychological side. They found that fear of a COVID-19 outbreak can have significant psychological repercussions on people, which requires more attention by the relevant authorities to cope with this impact. Also, the authors in Reference [40] had proposed a model that identified the risk of getting infected by tuberculosis based on several factors related to tuberculin skin, age, and weak immune system. They stated that those factors can increase the infection from 10% to 20%.

The authors in Reference [41] provided a model that predicts the course of the outbreak to help plan an efficient method of prevention. Model stages are SIDARTHE (susceptible, infected, diagnosed, ailing, recognized, threatened, healed, and extinct). It discriminates between infected people based on whether they have been diagnosed and on the severity of their symptoms. The simulation results obtained by combining the model with the available data on the COVID-19 pandemic in Italy indicate that it is an urgent necessity.

### 2.3. Healthcare Decision-Making

During the COVID-19 pandemic, the demand for emergency departments and medical equipment such as ventilators increased. Therefore, many studies have aimed to provide monitoring tools and models that help in making several medical decisions to mitigate potential risks, and these solutions include the following.

The authors in Reference [42] designed a prediction model called Conscious-based Susceptible-Exposed-Infective-Recovered (C-SEIR) model to ensure the usefulness of the lockdown and protective countermeasures in decreasing the influence of the pandemic in Wuhan city. The proposed model consisted of two classification groups, namely the quarantined suspected infection group (P), and the quarantined diagnosed infection group (Q), along with a blue/green curve with a solid line for daily patients and dashed line for cumulative patients. It showed that the result of the prediction is a double drop-down or increase based on the city lockdown precautions in Wuhan. The authors also gave guidance for protection against COVID-19, such as being educated about the virus, social distancing, and lockdown.

In Reference [43], the authors have developed a patient monitoring program that allows daily electronic checking of symptoms, providing advice and reminders via text messages, and providing care by phone. Patients registered in the system complete a daily questionnaire to evaluate 10 symptoms using a scale from 0 to 4. In addition to determining how much they feel the infection is affecting them, the number of analgesic/antipyretic tablets they take, and the temperature measured, questionnaire responses are used to classify patients and specify the care needed. The study focused on three measures, namely the number of patients monitored over time, the daily symptoms score, and daily ED referrals.

Likewise, the authors in Reference [44] developed a mobile app to track the spread of COVID-19 symptoms in the UK by analyzing a set of data reported by patients registered in the app, including location, age, health risk factors, symptoms, healthcare visits, and COVID-19 test results. Survey data helped in determining patients' type and intensity, availability of personal protective equipment, and work-related stress and anxiety.

The study presented in Reference [45] was concerned with evaluating one of the COVID-19 applications in terms of user satisfaction and the possibility of using the data collected to support decision-makers and healthcare providers. The app collects information daily from patients, including symptoms, vital signs, and an assessment of their satisfaction with the services provided by the app. The data collected is distributed on an interactive map according to the postal code for each user, which helps in knowing the regional distribution of the spread of infection in addition to the percentage of healthcare consumption in each region.

Another study [46] provided an analytical model for predicting patient census and estimating ventilator needs for a given hospital during the COVID-19 pandemic. Through this study, it was noticed that the estimation of the bed and ventilator needs is influenced by the length of hospital stay, and the number of days of inpatient ventilator use. Also, there was no relationship between the age of hospitalized patients and the likelihood of needing a ventilator, or between the inpatient gender and the length of stay. They recommended that each hospital relies on its internal data for accurate resource planning.

Furthermore, the Institute for Health Metrics and Evaluation (IHME) COVID-19 health service utilization forecasting team conducted a study to predict the expected daily use of health services and the number of deaths due to COVID-19 for the next four months from the date of the study for each state in the US [47].

The authors in Reference [48] tried to describe the clinical characteristics and identified factors that predict intensive care unit (ICU) admission for COVID-19 patients. They found that the need for a COVID-19 patient to enter the ICU can be predicted by checking a set of medical parameters that can be easily obtained: age, fever, and tachypnea with/without respiratory crackles. They used the EHRead [49] technique that was developed by Savana to extract information from the medical records. Also, deep learning convolutional neural network classification methods are used to classify the extracted data.

The authors in Reference [50] provided a data-driven framework to pre-assess the risks of the COVID-19 pandemic and to identify high-risk areas in Italy. The framework assesses the risk index using a function consisting of three criteria, namely disease risk, area exposure, and the vulnerability of its population. The twenty Italian regions are classified based on available historical data, which include population density, age, human mobility, air pollution, and winter temperature. The study showed a correlation between the risk index and the number of deaths, infected, and patients in ICU. They also provided a policy model to assist authorities in making several decisions.

Moreover, regional healthcare models have been developed to estimate the pandemic, like the simulation approach developed at the University of Pennsylvania called Monte-Carlo [51]. Such models can be used to manage facilities and plan for an anticipated increase in patient numbers, but not for an estimate of daily

operational needs. Applying the Pennsylvania model in an individual hospital requires unknown parameters like the proportion of the region's patients expected to visit that hospital, and the percentage of the regional population isolated sufficiently to avoid infection.

## References

1. Lee, I.-K.; Wang, C.-C.; Lin, M.-C.; Kung, C.-T.; Lan, K.-C.; Lee, C.-T. Effective Strategies to Prevent Coronavirus Disease-2019 (COVID-19) Outbreak in Hospital. *J. Hosp. Infect.* 2020, 105, 102–103.
2. Iacobucci, G. Covid-19: Emergency Departments Lack Proper Isolation Facilities, Senior Medic Warns. *BMJ* 2020, 368, m953.
3. Worldometers Coronavirus Cases. Available online: (accessed on 30 December 2020).
4. Da'Ar, O.B.; Haji, M.; Jradi, H. Coronavirus Disease 2019 (COVID -19): Potential Implications for Weak Health Systems and Conflict Zones in the Middle East and North Africa region. *Int. J. Health Plan. Manag.* 2020, 35, 1240–1245.
5. Ajah, I.A.; Nweke, H.F. Big Data and Business Analytics: Trends, Platforms, Success Factors and Applications. *Big Data Cogn. Comput.* 2019, 3, 32.
6. White, T. *Hadoop: The Definitive Guide*, 3rd ed.; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2012; ISBN 9781449311520.
7. Andreu-Perez, J.; Poon, C.C.Y.; Merrifield, R.D.; Wong, S.T.C.; Yang, G.-Z. Big Data for Health. *IEEE J. Biomed. Health Inform.* 2015, 19, 1193–1208.
8. Hagar, Y.; Albers, D.; Pivovarov, R.; Chase, H.S.; Dukic, V.; Elhadad, N. Survival Analysis with Electronic Health Record Data: Experiments with Chronic Kidney Disease. *Stat. Anal. Data Min. ASA Data Sci. J.* 2014, 7, 385–403.
9. Wang, Y.; Kung, L.; Wang, W.Y.C.; Cegielski, C.G. An Integrated Big Data Analytics-Enabled Transformation Model: Application to Health Care. *Inf. Manag.* 2018, 55, 64–79.
10. Wong, H.T.; Yin, Q.; Guo, Y.Q.; Murray, K.A.; Zhou, D.H.; Slade, D. Big Data as a New Approach in Emergency Medicine Research. *J. Acute Dis.* 2015, 4, 178–179.
11. Viceconti, M.; Hunter, P.; Hose, R. Big Data, Big Knowledge: Big Data for Personalized Healthcare. *IEEE J. Biomed. Health Inform.* 2015, 19, 1209–1215.
12. Erekson, E.A.; Iglesia, C.B. Improving Patient Outcomes in Gynecology: The Role of Large Data Registries and Big Data Analytics. *J. Minim. Invasive Gynecol.* 2015, 22, 1124–1129.

13. Nadkarni, G.N.; Coca, S.G.; Wyatt, C.M. Big Data in Nephrology: Promises and Pitfalls. *Kidney Int.* 2016, 90, 240–241.
14. Davis, R. Integrating Digital Technologies and Data-driven Telemedicine into Smart Healthcare during the COVID-19 Pandemic. *Am. J. Med. Res.* 2020, 7, 22.
15. Valenzuela, S. Unpacking the Use of Social Media for Protest Behavior. *Am. Behav. Sci.* 2013, 57, 920–942.
16. Sheng, J.; Amankwah-Amoah, J.; Wang, X.; Khan, Z. Managerial Responses to Online Reviews: A Text Analytics Approach. *Br. J. Manag.* 2019, 30, 315–327.
17. Rehman, A.; Naz, S.; Razzak, M.I. Writer Identification using Machine Learning Approaches: A Comprehensive Review. *Multimed. Tools Appl.* 2019, 78, 10889–10931.
18. Wang, Y.; Kung, L.; Byrd, T.A. Big Data Analytics: Understanding its Capabilities and Potential Benefits for Healthcare Organizations. *Technol. Forecast. Soc. Chang.* 2018, 126, 3–13.
19. Agbehadji, I.E.; Awuzie, B.O.; Ngowi, A.B.; Millham, R.C. Review of Big Data Analytics, Artificial Intelligence and Nature-Inspired Computing Models towards Accurate Detection of COVID-19 Pandemic Cases and Contact Tracing. *Int. J. Environ. Res. Public Health* 2020, 17, 5330.
20. Ponikowski, P.; Anker, S.D.; Alhabib, K.F.; Cowie, M.R.; Force, T.L.; Hu, S.; Jaarsma, T.; Krum, H.; Rastogi, V.; Rohde, L.E.; et al. Heart Failure: Preventing Disease and Death Worldwide. *ESC Hear. Fail.* 2014, 1, 4–25.
21. Vaishya, R.; Javaid, M.; Khan, I.H.; Haleem, A. Artificial Intelligence (AI) Applications for COVID-19 Pandemic. *Diabetes Metab. Syndr. Clin. Res. Rev.* 2020, 14, 337–339.
22. PEX Process Excellence Network 6 Ways Pharmaceutical Companies are Using Big Data to Drive Innovation & Value. Available online: (accessed on 28 December 2020).
23. Abdel-Basst, M.; Mohamed, R.; Elhoseny, M. A Model for the Effective COVID-19 Identification in Uncertainty enVironment using Primary Symptoms and CT Scans. *Heath Inform. J.* 2020, 1–18.
24. Stojanovic, R.; Skraba, A.; Lutovac, B. A Headset Like Wearable Device to Track COVID-19 Symptoms. In Proceedings of the 2020 9th Mediterranean Conference on Embedded Computing (MECO), Budva, Montenegro, 8–11 July 2020; pp. 1–4.
25. Gordon, W.J.; Henderson, D.; DeSharone, A.; Fisher, H.N.; Judge, J.; Levine, D.M.; MacLean, L.; Sousa, D.; Su, M.Y.; Boxer, R. Remote Patient Monitoring Program for Hospital Discharged COVID-19 Patients. *Appl. Clin. Inform.* 2020, 11, 792–801.
26. Mishra, T.; Wang, M.; Metwally, A.A.; Bogu, G.K.; Brooks, A.W.; Bahmani, A.; Alavi, A.; Celli, A.; Higgs, E.; Dagan-Rosenfeld, O.; et al. Pre-Symptomatic Detection of COVID-19 from Smartwatch Data. *Nat. Biomed. Eng.* 2020, 4, 1208–1220.



27. Lan, F.-Y.; Filler, R.; Mathew, S.; Buley, J.; Iliaki, E.; Bruno-Murtha, L.A.; Osgood, R.; Christophi, C.A.; Fernandez-Montero, A.; Kales, S.N. COVID-19 Symptoms Predictive of Healthcare Workers' SARS-CoV-2 PCR Results. *PLoS ONE* 2020, 15, e0235460.
28. Richardson, S.; Hirsch, J.S.; Narasimhan, M.; Crawford, J.M.; McGinn, T.; Davidson, K.W.; The Northwell COVID-19 Research Consortium. Presenting Characteristics, Comorbidities, and Outcomes among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA* 2020, 323, 2052–2059.
29. Brown, C.; Chauhan, J.; Grammenos, A.; Han, J.; Hasthanasombat, A.; Spathis, D.; Xia, T.; Cicuta, P.; Mascolo, C. Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining 2020. Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data (ACM)*, New York, NY, USA, 25–27 August 2020; pp. 3474–3484.
30. Jeong, H.; Rogers, J.A.; Xu, S. Continuous on-Body Sensing for the COVID-19 Pandemic: Gaps and Opportunities. *Sci. Adv.* 2020, 6, eabd4794.
31. Alsofayan, Y.M.; Althunayyan, S.M.; Khan, A.A.; Hakawi, A.M.; Assiri, A.M. Clinical Characteristics of COVID-19 in SAUDI Arabia: A National Retrospective Study. *J. Infect. Public Heal.* 2020, 13, 920–925.
32. Khan, M.A.; Karim, R.; Kim, Y. A Two-Stage Big Data Analytics Framework with Real World Applications Using Spark Machine Learning and Long Short-Term Memory Network. *Symmetry* 2018, 10, 485.
33. Riaz, H.; Park, J.; Choi, H.; Kim, H.; Kim, J. Deep and Densely Connected Networks for Classification of Diabetic Retinopathy. *Diagn.* 2020, 10, 24.
34. Xiong, S.; Liu, L.; Lin, F.; Shi, J.; Han, L.; Liu, H.; He, L.; Jiang, Q.; Wang, Z.; Fu, W.; et al. Clinical Characteristics of 116 Hospitalized Patients with COVID-19 in Wuhan, China: A single-Centered, Retrospective, Observational Study. *BMC Infect. Dis.* 2020, 20, 1–11.
35. Chen, Q.; Xu, L.; Dai, Y.; Ling, Y.; Mao, J.; Qian, J.; Zhu, W.; Di, W.; Ge, J. Cardiovascular Manifestations in Severe and Critical Patients with COVID -19. *Clin. Cardiol.* 2020, 43, 796–802.
36. Sharp, A.L.; Huang, B.Z.; Broder, B.; Smith, M.; Yuen, G.; Subject, C.; Nau, C.; Creekmur, B.; Tartof, S.; Gould, M.K. Identifying Patients with Symptoms Suspicious for COVID-19 at Elevated Risk of Adverse Events: The COVAS Score. *Am. J. Emerg. Med.* 2020.
37. Benreguia, B.; Moumen, H.; Merzoug, M.A. Tracking COVID-19 by Tracking Infectious Trajectories. *IEEE Access* 2020, 8, 145242–145255.
38. Xie, C.; Zhao, H.; Li, K.; Zhang, Z.; Lu, X.; Peng, H.; Wang, D.; Chen, J.; Zhang, X.; Wu, D.; et al. The Evidence of Indirect Transmission of SARS-CoV-2 Reported in Guangzhou, China. *BMC*

Public Health 2020, 20, 1–9.

39. Khan, A.H.; Sultana, M.S.; Hossain, S.; Hasan, M.T.; Ahmed, H.U.; Sikder, T. The Impact of COVID-19 Pandemic on Mental Health & Wellbeing Among Home-Quarantined Bangladeshi Students: A Cross-Sectional Pilot Study. *J. Affect. Disord.* 2020, 277, 121–128.
40. Horsburgh, C.R. Priorities for the Treatment of Latent Tuberculosis Infection in the United States. *N. Engl. J. Med.* 2004, 350, 2060–2067.
41. Giordano, G.; Blanchini, F.; Bruno, R.; Colaneri, P.; Di Filippo, A.; Di Matteo, A.; Colaneri, M. Modelling the COVID-19 Epidemic and Implementation of Population-Wide Interventions in Italy. *Nat. Med.* 2020, 26, 855–860.
42. Chen, B.; Shi, M.; Ni, X.; Ruan, L.; Jiang, H.; Yao, H.; Wang, M.; Song, Z.; Zhou, Q.; Ge, T. Visual Data Analysis and Simulation Prediction for COVID-19. *IJEE* 2020, 6, 95–114.
43. Kricke, G.E.; Roemer, P.; Barnard, C.; Peipert, J.D.; Henschen, B.L.A.; Bierman, J.; Blahnik, D.; Grant, M.; Linder, J.A. Rapid Implementation of an Outpatient Covid-19 Monitoring Program. *NEJM Catal. Innov. Care Deliv.* 2020, 1.
44. Drew, D.A.; Nguyen, L.H.; Steves, C.J.; Menni, C.; Freydin, M.; Varsavsky, T.; Sudre, C.H.; Cardoso, M.J.; Ourselin, S.; Wolf, J.; et al. Rapid Implementation of Mobile Technology for Real-Time Epidemiology of COVID-19. *Science* 2020, 368, 1362–1367.
45. Timmers, T.; Janssen, L.; Stohr, J.; Murk, J.L.; Berrevoets, M. Using eHealth to Support COVID-19 Education, Self-Assessment, and Symptom Monitoring in the Netherlands: Observational Study. *JMIR mHealth uHealth* 2020, 8, e19822.
46. Epstein, R.H.; Dexter, F. A Predictive Model for Patient Census and Ventilator Requirements at Individual Hospitals During the Coronavirus Disease 2019 (COVID-19) Pandemic: A Preliminary Technical Report. *Cureus* 2020, 12, e8501.
47. IHME. COVID-19 Health Service Utilization Forecasting Team Forecasting COVID-19 Impact on Hospital Bed-Days, ICU-Days, Ventilator-Days and Deaths by US State in the Next 4 Months [PRE-PRINT]. *Medrxiv* 2020.
48. Izquierdo, J.L.; Ancochea, J.; Soriano, J.B. Savana COVID-19 Research Group Clinical Characteristics and Prognostic Factors for Intensive Care Unit Admission of Patients With COVID-19: Retrospective Study Using Machine Learning and Natural Language Processing. *J. Med. Internet Res.* 2020, 22, e21801.
49. Medrano, I.H.; Guijarro, J.T.; Belda, C.; Ureña, A.; Salcedo, I.; Espinosa-Anke, L.; Saggion, H. Savana. Re-using Electronic Health Records with Artificial Intelligence. *Int. J. Interact. Multimed. Artif. Intell.* 2018, 4, 1.

50. Pluchino, A.; Biondo, A.E.; Giuffrida, N.; Inturri, G.; Latora, V.; Moli, R.L.; Rapisarda, A.; Russo, G.; Zappala', C. A Novel Methodology for Epidemic Risk Assessment: The case of COVID-19 outbreak in Italy. *Arx. Prepr. Arx.* 2020, 2004, 1–37.
51. Weissman, G.E.; Crane-Droesch, A.; Chivers, C.; Luong, T.; Hanish, A.; Levy, M.Z.; Lubken, J.; Becker, M.; Draugelis, M.E.; Anesi, G.L.; et al. Locally Informed Simulation to Predict Hospital Capacity Needs During the COVID-19 Pandemic. *Ann. Intern. Med.* 2020, 173, 21–28.

---

Retrieved from <https://encyclopedia.pub/entry/history/show/22486>