



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

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

CHAPTER 2

Investigation of COVID-19 and scientific analysis big data analytics with the help of machine learning

Victor Chang^c, Mohamed Aleem Ali^a, and Alamgir Hossain^b

^aCybersecurity, Information Systems and AI Research Group, School of Computing, Engineering and Digital Technologies, Teesside University, Middlesbrough, United Kingdom

^bAcademic Affairs & Research, Cambodia University of Technology and Science (CamTECH), Phnom Penh, Cambodia

^cDepartment of Operations and Information Management, Aston Business School, Aston University, Birmingham, United Kingdom

1 Introduction and background

This project is based on the investigation process of COVID-19 and scientific analysis, big data with the assistance of machine learning. The extended COVID-19 is a global epidemic that has produced an exponentially rising data value that may be attached to enhance the considerate research of big data management. Additionally, to demonstrate the importance of policymakers, scholars, and practitioners for a deeper and better understanding of the assortment of investigative tools that might be exploited to respond better and anticipate unexpected risks and events. Deal with the global pandemic, scientists, clinicians, and WHO (World Health Organization) are investigating or searching the new technologies to analyze an infected person in different stages, to control the virus spread, develop the vaccine for a remedial infected person. Current investigation recognized that artificial intelligence and machine learning are capable technology engaged with the different providers of health care as they consequence the better power of processing and obtain a better result with accurate human care work. This project will explain the solving problems of COVID-19 with the benefit of machine learning. The objectives will describe the methods, techniques, and accurate process of data analysis with a different algorithm utilized in machine learning. The main focus of this project is on the pandemic of COVID-19 and how the technologies of machine learning are currently engaged to resolve the disputes during the situation of COVID-19. See Fig. 1.

Machine Learning Applied to Big Data

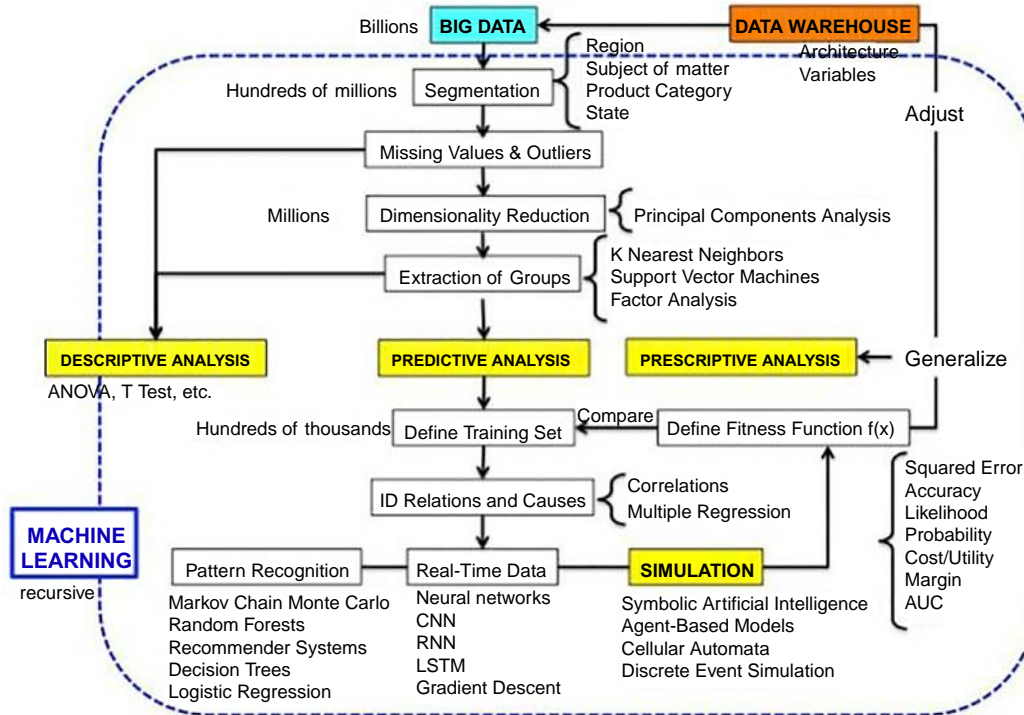


Fig. 1 Machine learning in big data. ([https://analytixon.com/images/.](https://analytixon.com/images/))

“2019-nCoV” or “the novel coronavirus” or “COVID-19” as given the same by the world health organization (WHO), is the responsible organization regarding the present outburst of pneumonia, the commencement of which was in December 2019 in the urban of Wuhan, China. This virus is pathogenic. The reservoir of occurrence of this virus are bats, but the host that served as an intermediary has still been undetected. Though, there is ongoing work in china in three very significant areas regarding the advising on the awareness of the pathogenic origination of the outbreak. This involves inquiries of the cases who are showing some symptoms, which were occurring in the city of Wuhan around 2019, December, collection of ecological samples from the market of wholesale seafood, and also additional markets within such areas, the collection regarding the reports made on the detailing of the origination, the kinds of wildlife species which were marketed in those markets, and what are the destination of such animals once there has been closing of the markets. Mostly the gastrointestinal and the respiratory tract infections are caused by the coronavirus and they are categorized and four significant kinds: Delta coronavirus, Gamma coronavirus, and Alpha coronavirus and Beta coronavirus. Infection is caused to birds from the first two kinds of the coronavirus, while mammals are infected from the last two. When it comes to human beings, coronavirus has been formally categorized into six kinds. The Middle East Respiratory Syndrome coronavirus (MERS-CoV), HCoV-OC43, HCoVHKU1, Severe Acute Respiratory Syndrome coronavirus (SARS-CoV), which is the kind of the Beta coronavirus, HCoV229E as well as HCoV-NL63, which are the associate of the Alpha coronavirus. Coronavirus was not considered a global concern until 2003 when the pandemic of SARS arrived, which first-order continued in the year 2012 by MERS, and then with the most recent cases and outbreaks by the COVID-19. All these three viruses are highly pathogenic. Their spreading is from bats to palm civets, and then in the end to the humans. Suppose there will be unsafe touching between the infected person and the person who can be infected, COVID-19 splendid given by dust particles and the fomites. Although based on the empirical evidence, transmission through the air is not recorded for coronavirus, and it is not considered a very important transmission engine. However, there is a possibility that if such particles, which are aerosol-generating, will be carried out in the medical facilities, they can become an important mode of transmission. In some patients, it was observed that it could also be spread fecal spreading. Also, the fecal and oral routes are not quite identified much significantly. Their functions and the importance of spreading COVID-19 are an ongoing process and still need to be determined. In the

second week of April 2020, approximately 1,873,858 laboratory-confirmed cases were recorded. After further studying the age groups, it was determinable that a maximum number of confirmed cases bar between the age groups of 30–69. In all the recorded cases, approximately 21.6% of the cases were found in the working class or employees by profession and farmers. The 2019 novel coronavirus, or acute respiratory syndrome known nowadays, is rapidly spreading from China to the rest of the world. Around 95,000 cases of the disease in 2019 and 34,000 death have been recorded. This is fortunate that the children are not affected by the virus with no deaths. But the cause of the future is unknown.

However, currently there is too much to be concerned about in relation to the latest coronavirus. However, it can be said that the humans were infected because of transmission through animals. Therefore, it becomes necessary to determine particular animals and various other sources, the line of the broadcast, the sequence of incubation, characteristics of the susceptible public, and survival rate. However, there is not much clinical knowledge of the COVID-19 and other details regarding it. Like the outbreak curve, the spectroscopy of its virility, the pathogenesis of its dissemination, the observations of the autopsy, and any types of clinical responses. COVID-19 is a lot different than other viruses that have emerged in history. The reach of COVID-19 has been to almost 150 nations and the country of China. The outbreak of this virus was professed as a worldwide epidemic, according to the world health organization. In a very short span of time and only in the second week of April 2020, the total number of cases of COVID-19 exceeded 1,873,858 and the deaths were recorded of 1,160,450 worldwide. The global epicenter of the coronavirus was the United States of America. Though the outbreak started in China, more than one-third of the coronavirus cases were found outside the country of China. It was very difficult to establish any kind of connection between the pandemics that occurred in the past and this one, for example, swine flu, bird flu and more so. Since it was so new and no vaccine could be made in such a short period, the only way left to fight this virus was some precautions for its prevention. Some kinds of preventive measures and policies were advised and enforced by the world health organization. Since this epidemic is continuous and increasing rapidly, some precautions are necessary to be adopted, such that the risks of its spreading could be significantly reduced and its transmission chain could be broken. The advice given by the world health organization is:

Regular washing of hands with any hand wash that is alcohol-based or soap and water is helpful.

Proper distance maintenance.

There is no touching of nose, head, and ears.

It requires a Proper covering of the nose and mouth to prevent the spread of viruses in the air.

In case any person has fatigue or faces any trouble in breathing, it should be taken seriously and provided with early medical attention.

Take proper precautions, like isolation, if recently any place has been visited by you where there are high chances of spreading coronavirus.

Since a progressive “datafication” has been witnessed in the social life. There has been monitoring of human activities as well as the interactions of humans with the environment, with effectiveness such that a huge amount of digital footprints can be generated. Big data examination is akin to a very complicated procedure of examining data to uncover important information. For example, correlations, the customers’ preferences, many hidden patterns, and many kinds of market trends will be helpful for the organization in making correct and informed decisions.

On a wider scale, the technologies and techniques of data analytics are very significant for the organization since they provide a way in which sets of data could be analyzed and new information could be gathered. Basic questions queries regarding the performance and operations of the business might be answered with business intelligence support. Big data analytics can also be referred to as advanced analytics, which includes complicated applications with traits like predictive models, the algorithms of statistics and analytics system-powered what-if analysis.

Since the collection, processing, and cleaning of the data are done, the data will be analyzed with analytics software. This involves the tools for the mining of data, which sifts through the sets of data in search of establishing or determining any kinds of patterns or relationships. Text mining, predictive analysis-helps build software that can forecast the customer’s behaviors and other developments related to the future. Artificial intelligence and visualization tools for data science and deep learning can also be termed the advanced offshoot of data-oriented machine learning. Some examples of key big data analytics tools which are put into use for the enabling of big data analytics processes involve:

- Hadoop: is an open basis outline for the storage and dispensation of big data sets. Hadoop can handle huge amounts of organized as well as

formless data. According to Hadoop, the advantages include the following characteristics.

Performance, scalability, data locality, support for multiple languages, open source, schema independent, fault tolerant, share-nothing architecture, cost-effectiveness, high-throughput and low latency, with explanations as follows.

- Predictive analysis: predictive analysis hardware as well as software process a huge amount of complicated data and make use of statistical algorithms and machine learning techniques so that predictions regarding the outcomes of future events could be made. Organizations use predictive analytical tools for the detection of fraud, marketing, and risk assessment.
- Stream analytics tools are used to filtrate, aggregate, and analyze big data that may stow in various arrangements or on different stages.
- Distributed storage data: the replication of which is done on non-relational databases. Its use could be made to a degree in contradiction of the bulge failures, which are independent, data that is not apt or corrupted.
- NoSQL databases: these are useful when huge sets of distributed data are used, and these are non-relational data management systems. A fixed scheme is not used or required by them, and this feature makes them ideal for the data, which is raw and unstructured.
- Data Lake is like a repository that helps store enormous amounts of data unruffled by various sources. The data which the data warehouses typically store is done by using predefined schemas.
- Big data mining tools: the enabled business in mining huge amounts of data is structured and unstructured data.
- In-memory data fabric: it has a distribution of huge amounts of data across the systems of mineral resources. It helps in hiding low latency for the data access as well as processing.
- Data integration software has enabled the big data to be rationalized on various stands, including Hadoop, Amazon EMR, Mongo DB, and apache.
- Spark: it is a cluster of open-source computing frameworks used for the batch and the streaming of data processing.
- Data quality software: cleanses as well as provides enrichment of the big data sets.

Machine knowledge is a way of data analysis due to which analytical model structure is computerized. It is like a division of artificial intelligence based on the idea start a lot could be learned by the systems from the data, like

identifying patterns and making decisions with the minimum kind of human interruption. Due to the revolution of new computing technologies, machine learning in the present scenario does not write machine learning in the past. Its origination was recognizing pattern yes theory that the computers could learn without programming for the performance of those particular tasks, so many researchers were interested in the concept of artificial intelligence wanted to observe is learning from data could be done by computers.

The repetitive aspect in machine learning is very significant since models have to face the risk of exposure to new data. They have the ability to adapt independently. Learning is done from the competitions that were previously performed to produce reliable, repeatable decisions and results. This is a science that is not new, but it has been considered change and has definitely gained new and fresh momentum.

Recommendations regarding online offers like those from Amazon and Netflix can be said to be machine learning applications put into use in everyday life. Understanding and knowing what is being said by the customers about you and your organization on media platforms like Twitter is a classic example of machine learning combined with linguist rule creation. Detection of fraud is one of the most obvious and significant used in today's world.

The reasons for researching interest in machine learning are that some factors obligate Bayesian analysis and data mining much more famous than they were constantly. There are also other concerns, such as increasing volumes and the cost of computational processing. Users need a much lower and comparatively powerful and affordable data storage. In summary, it has become reasonable to ghastly and repeatedly yields the models that can examine bigger, more complicated data. The delivery will be much faster and the results will be more accurate. And this all will be done even if the scale of data is huge. Once precise models are built, the organization automatically has a much better chance of identifying profitable opportunities and avoiding unknown risks.

Many industries whose daily work is to go through huge quantities of data have understood the importance of machine learning technology. Through getting detailed visions from such usage data, most often in the actual time, establishments are now capable of working more efficiently and gaining an advantage over their competitors. Organizations provide financial services such as banks and many government agencies like public safety, healthcare industries, so many websites, organizations that are engaged in building new energy sources, and transportation industries—they can generate abundant data sources.

Machine learning depends upon CT analysis, which is also suggested as the promising screening tool for COVID-19. There is one more PCR test that is needed to do on the COVID-19 patient—the result to be cautiously understood. During the pandemic, the studies done of hamper inherently by the disease occurrence and the nature with the selected of the members, evaluation of CT. See Fig. 2.

The procedure also needs to be skilled on the variety of diseases, counting early-stage cases and asymptomatic. The elucidation of CT machine learning will be functional in real-world data with poise. Analytics of the big data calculator mortality risk from the COVID-19: the pandemic has warned the people, health officials searching continuously for the different tool to help the resources and give guidance in decision-making. There is a squad from John Hopkins school of public health purchase Big data analytics to build COVID-19 risk calculator, which well informed the public health policy around the resources of prevention like face mask (N95).

Artificial intelligence shows that the vaccine of the COVID-19 is less effective. With the use of authorization, the FDA recently granted the emergency for the new vaccines of the COVID-19 Hindi looking forward to the beginning to the end of the pandemic. The researcher news the tool to investigate the vaccine, which is similar to the COVID-19 vaccine and found it less effective in Asian ancestry or black people. The result highlights the ethnic disparities and stark racial, which have been highlighted consistently throughout the pandemic. Data set of the Google map will be freely available to a researcher: the new ways parked with the name of the COVID-19 of sharing the data and access in the healthcare. Google announced in late March that it would permit access to the critical coronavirus data to the researcher through its COVID-19 public data set program, which has the objective to boost the solution during the global pandemic.

Using machine learning to discover the COVID-19 solution by the data: India is at a pandemic stage. The researcher is working to find the therapies for COVID-19 which is potential by using the machine learning tool and artificial intelligence. To graduates from the institute of data science at Columbia University build a startup called EVQLV, which made the algorithm capable of screening, optimizing and computationally generating millions of antibodies. Find the COVID-19 patients who are at high risk. In

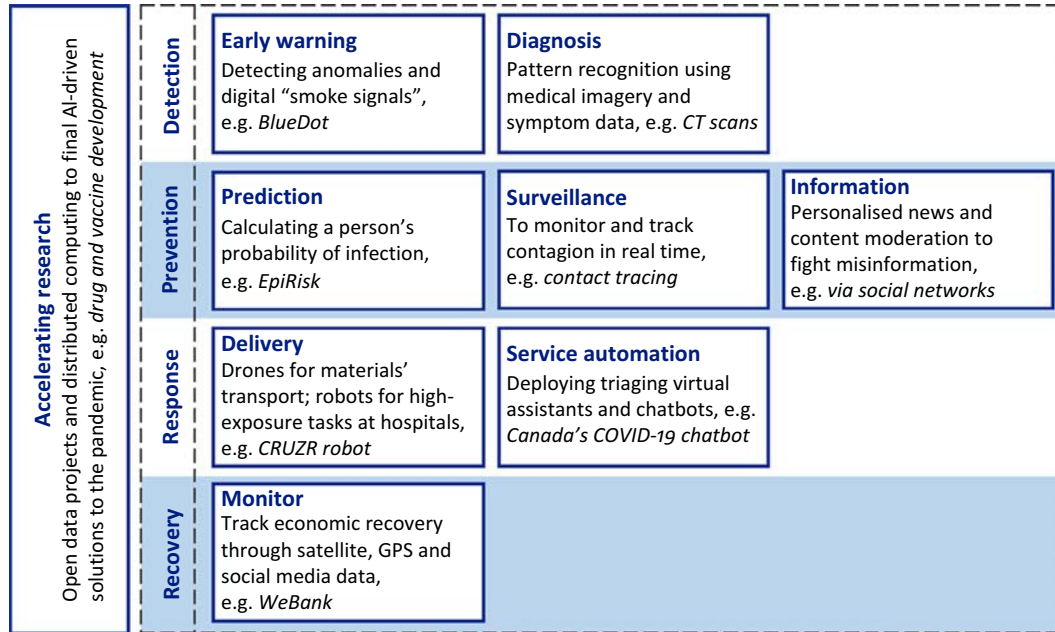


Fig. 2 Use of machine learning in COVID-19. (Data from *Nirmala, A. P., & More, S. (2020, December). Role of Artificial Intelligence in fighting against COVID-19. In 2020 IEEE International Conference on Advances and Developments in Electrical and Electronics Engineering (ICADEE) (pp. 1-5). IEEE.*)

The beginning, the virus spreading in the whole us, the population was at a higher risk while spreading this virus.

The virus has impacted the people or the elderly with good condition and affects the population who are the minorities and individuals of the lower status.

Accepting the COVID-19 as a big data analytic issue: the data has been considered as the midpoint of the pandemic Investigation effort. The extent of the coronavirus has resulted in its intricate nature and requires assessors and analysis of the massive amount of information. The particular interview has determined that there are different ways in which resources and development can use machine learning, natural language, artificial intelligence, and dispensation to track, understand, and contain coronavirus. The institute has been providing the government entities industries and Research Organization exist and innovation in artificial intelligence tools. Data platform helps track Hospital bed capacity during the outbreak of COVID-19: the spreading of COVID-19 has meant that hospitals are required to prepare for the poorest. In a particular system that has previously been tense, there are probable waves that result in highly contagious patients and can result in disaster translation. With the usage of Big Data Analytics, the organization can monitor and track the use of critical resources. Occasionally, the speed and accuracy of machine learning algorithms are to be stated on the foundation of the presentation in scenarios that are not capable of reflecting the actual clinical practice. Sometimes, the comparisons made between the algorithms and the preferences of humans are very imbalanced. In many cases, a computer has been trained to detect the particular abnormality related to parenchymal disease related to COVID-19. In contrast, any abnormality which is related to the finding of the pulmonary nodules becomes the responsibility of the radiologist to detect the same. There is no doubt that many exciting opportunities were offered by the COVID-19 pandemic regarding artificial intelligence research. With all the understanding which could be developed regarding the coronavirus, it is very easily determinable that this virus progresses when it reaches different stages. Regarding the previous pandemics that have occurred and their in depth-details, it can be said that this pandemic is not very different from the others. With the help of the correct collaboration between the clinical and machine learning know-how, the present public health crisis strength marks the starting of the decade when artificial intelligence regarding health care will deliver according to its promises of transformative clinical impacts.

2 Literature review

As per the words of [Sheng, Amankwah-Amoah, Khan, and Wang \(2020\)](#), COVID-19 has become a worldwide health disaster, approximately 16 million persons are infected by this virus and more than 666,000 deaths are faced till the 29th of July 2020 all over the world. This resulted in excessive demand for resources in several countries to stop the spread of the COVID-19 pandemic. COVID-19 disease confirmed its first case in December 2019 in China. The health crisis is demanding effective vaccines and drugs for treatment processes and mitigating the spread of infection. The majority of the contaminated areas follow lockdown policies, quarantine, and social distancing measures because of the limited validated therapeutics to reduce the spread of infection. These measures are proven helpful in slowing down the spread of infection, but it is not a permanent solution to eliminate the infection. The fear of spreading infection into the second wave was increasing. To limit the second potential outbreak, several measures like identifying hotspots, contact tracing are required.

Inconsistency has been observed in the symptoms of the COVID-19; the symptoms range from flu to fulminant pneumonia and serious respiratory distress syndrome. Various techniques such as big data analytics and machine learning investigate the information regarding the COVID-19 and scientific analysis. Machine learning is used to harness, predict, understand, and respond to be taken for future events.

As per the words of [Lalmuanawma, Hussain, and Chhakchhuak \(2020\)](#), numerous health care experts, clinicians, and scientists worldwide are working on researching new and suitable technologies for supporting the handling of the COVID-19 pandemic. Previous application of Machine Learning and Artificial Intelligence on this epidemic situation encouraged the investigators by providing a novel approach for battling against the coronavirus outbreak. Artificial intelligence, besides machine learning, provides a vital approach in screening, forecasting, predicting, and developing a drug for SARS-CoV-2 and contact tracing. These approaches helped the researchers by enlarging various angles towards the research. WHO and co-operating clinicians and several national authorities all over the world are fighting the pandemic? WHO avowed the COVID-19 situation as a global well-being emergency. Several researchers and scientists are searching for technology for screening the virus-infected patients, finding an appropriate clinical trial, developing a cure vaccine, controlling virus spread, and tracing the contact of virus-infected patients. Several healthcare providers

hire machine learning and artificial intelligence because they result in reliable, speedy processing power, better scaling, and outperforming humans' in particular healthcare tasks. Most of the clinician and health care industries employed several AI and ML technology for handling the pandemic and addressing the challenges at the time of the outbreak.

AI in the industry of medicine is used for supporting the decisions of the clinician. AI is used for selecting a supplementary medication for investigating about proposed SARS-CoV-2. There are several technologies and methodologies are used for classifying and addressing the classical methods based on statistics to advanced and modern ML and AI algorithms. These technologies are promising improvements in predicting, screening, forecasting, and tracing contact and developing drugs. They help predict incidents and the number of deaths due to pandemics. They are used to perform predictive analysis, diagnostic and descriptive analytics, and build complex models for enlarging the datasets to predict future outcomes (Corsi, de Souza, Pagani, & Kovaleski, 2020).

In the early period of the spread and outbreak of COVID-19, it was necessary to adopt scientific data analysis methods for making effective decisions regarding the prevention of the COVID-19 situation. With the increasing popularity of technology on the internet, people around the globe can easily interact, communicate, learn, shop and others. See Fig. 3 for details.

Hence, using appropriate technology for analyzing big amounts of data associated with information networks, accurate predictions can be made on outbreak and blowout of COVID-19, way and cause of spread, and the next stage of spreading can formulate policy measures accumulated with it. Controlling and preventing these types of situations are classified in different stages: monitoring of the epidemic situation, early warning, and prevention of disaster, disaster recovery, and relief.

Machine learning and AI technology are useful in speeding up processing power, scaling up, and outperforming humans in specific healthcare tasks. Various healthcare clinicians and industries employ the technology of ML and AL for tackling up the situation of the COVID-19 epidemic for addressing challenges during the time of the outbreak. Even if AI cannot replace human interaction in healthcare, it can support decisions (Syeda et al., 2021).

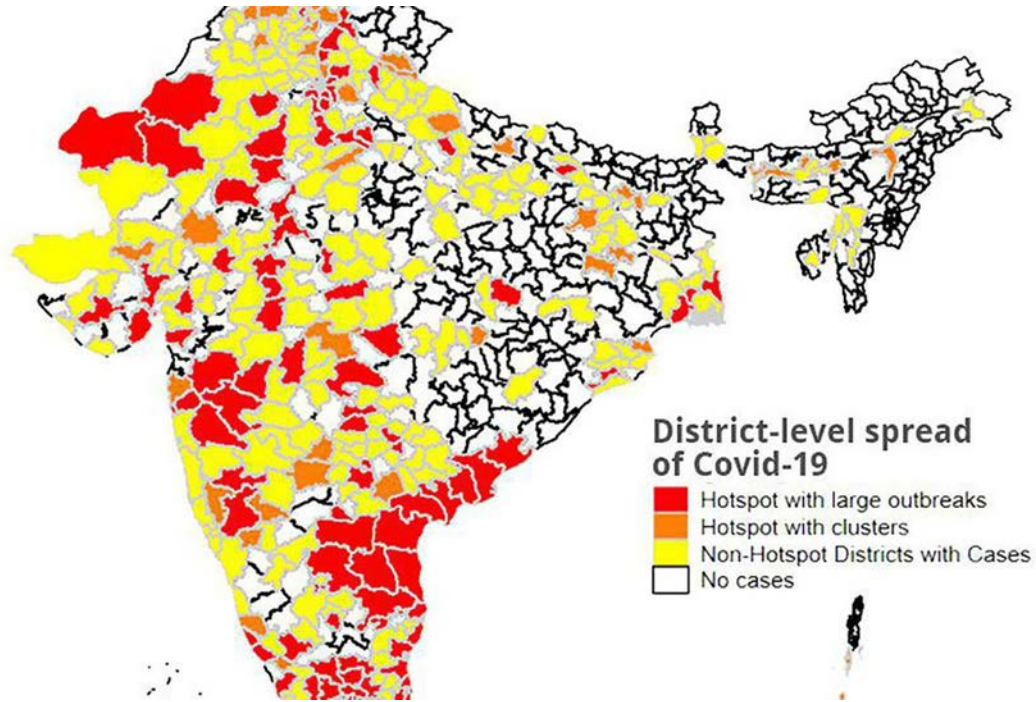


Fig. 3 COVID-19 statistics. (Source india.gov.in; <https://scroll.in/author/18857>.)

3 COVID-19 pandemic in the new era of big data analytics: Methodological innovations and future research directions

As per [Jia, Guo, Wang, and Barnes \(2020\)](#), the outbreak of the COVID-19 has generated the epidemic globally with an extraordinary volume of data and exponential amounting, which many enhance the empathy of the data administration research. The amount of data can be very huge and a symbolized need among the toppers policymakers and practitioners deeper and better understanding of the range for the analytical tool that could be used to respond better unanticipated to search unexpected Black swan risk and event. See [Fig. 4](#).

Big data is specified by volume, which has great variety, complexity and velocity. The big data benefit does not seem to result from the size of data but only appears when functioning with the fine-grained data. It is noticed that the many studies joined many sources of the information, which enables 360 points of degrees the study of the object. This might involve the big structured data sets of unstructured data such as web data, tweets, text, news-feed social media, and social bots. Diversification of the information needs a hybrid of two methods or more than two methods to analyze and process the information. The characteristic example of using machine methods is determining the formless data and involving them as inputs into models and statistics.

As per [Haleem, Javaid, Khan, and Vaishya \(2020\)](#), data sources are compulsory in relation to enhancing predictive performance. Considering the complexity and variability of the big data, studies in the area have the tendency to compare the output of the many algorithms or joined a few algorithms to optimize the predictive power. By choosing the algorithm which is having the best performance is not straightforward. It has been determined that no particular can be fit cases. It is being suggested that a proper model should be selected that offers the best balance between variance and bias ([Agbehadji, Awuzie, Ngowi, & Millham, 2020](#)).

3.1 Deep learning applications for COVID-19

The definition of which technique of detailed learning is the use of network switch is neural and one or more than two layers. Network, a neural layer, is a typical “layer” made of the parametric transformation of a nonlinear input. Moreover, it has been determined the deep learning is considered an

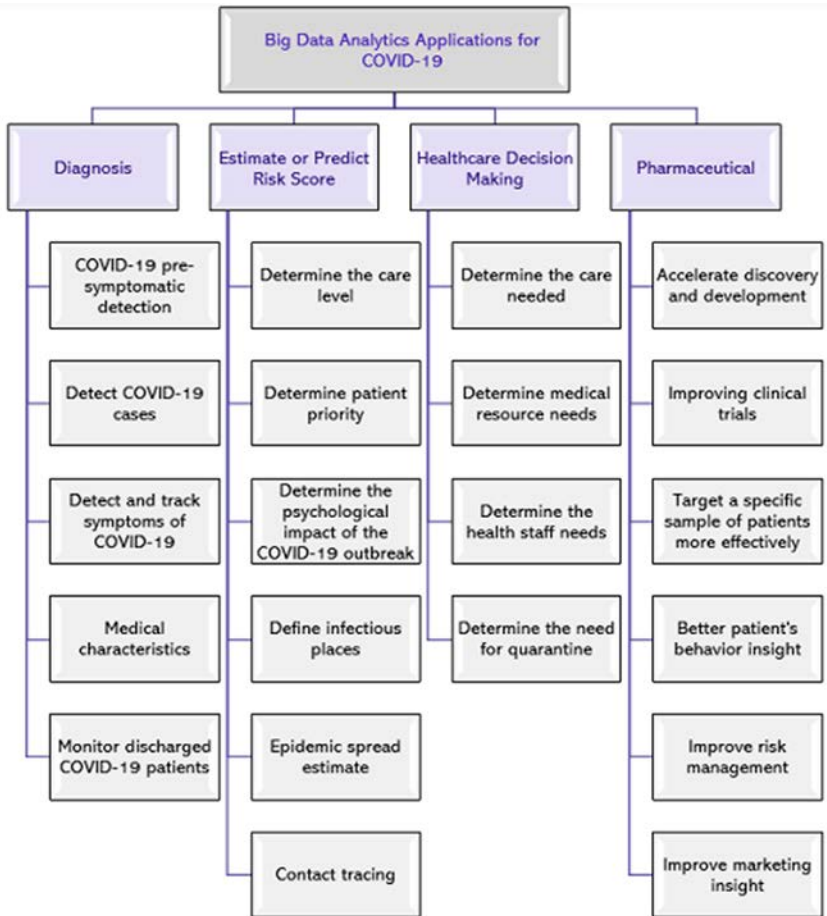


Fig. 4 Big data analytics in COVID-19.

important application in COVID-19, including misinformation detection preparation mining and public sentiment analysis. Computer vision is also considered an important application to the domain in relation to deep learning. The solicitation of deep learning can help in fighting the promising coronavirus, but it is considered quite significant to be acquainted in relation to the disadvantages of deep learning. The main issues are related to data privacy, interpretability, learning from the limited labeled datasets and generalization matrix. It is only analyzed and is quite crucial in interpreting the output of deep learning models. See Fig. 5.

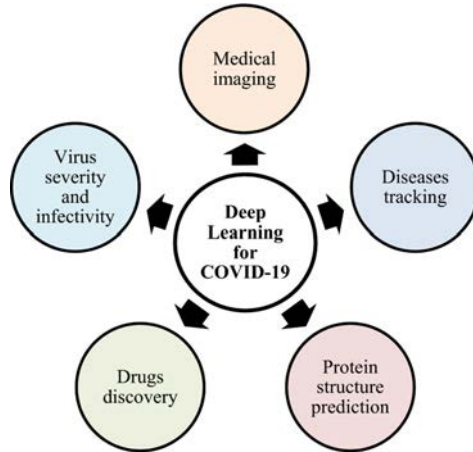


Fig. 5 Deep learning as a tool in COVID-19. (Data from Asraf, A., Islam, M. Z., Haque, M. R., & Islam, M. M. (2020). Deep learning applications to combat novel coronavirus (COVID-19) pandemic. *SN Computer Science*, 1(6), 1–7.)

3.2 Big data analytics as a tool for fighting pandemics: A systematic review of literature

The big data term has become extensively used in the growth of data exponentially in recent years. Because of this popularity, the meaning of this is diverse. It can also understand as the volume of data which is huge with the diversity of topics and with the highest speed and it will also in the variety which is coming from many different sources like sensors, social networks and the cell phones and many other devices which will need the specific tools and techniques for their process.

As shown in Fig. 6 big data science study has many different aspects of big data that involve data storage capacity and analysis of the speed. Show the study of the different areas it can be employed, such as social media, commercial activity and health data. This study has the common thing, which is needed for the analysis of the data symbol analysis will not have the configuration or the capacity which is the same that the big data analytics will provide to the users (Sheng, Amankwah-Amoah, Khan, & Wang, 2020). BDA is becoming the modern practice for data analyzing also find as the analysis, which is specific for this scenario.



Fig. 6 Big data analytics in COVID-19. (Data from Jia, Q., Guo, Y., Wang, G., & Barnes, S. J. (2020). *Big data analytics in the fight against major public health incidents (including COVID-19): A conceptual framework*. International Journal of Environmental Research and Public Health, 17(17), 6161.)

4 Review of big data analytics, artificial intelligence and nature-inspired computing models towards accurate detection of COVID-19 pandemic cases and contact tracing

Contact tracing with big data analytics: the contact drawing through Big Data Analytics is considered as a significant domain that can result in assisting the outbreak of the disease using different sources of data that includes tags used in social media, Post with metadata, list of passengers, the logo of a means of transportation that people travel with smart cards to metro and the usage of different credit cards are all considered as appreciated sources of data.

In this article, the contact tracing model has been estimated that results in the purpose of the rate of the broadcast of infection. This model can be easily identified that the potentially infected individuals before the emergence of the system. It has been determined that the most distinguished feature of humanoid social contact is that it includes the use inconsistency in the degree and number of associates in both the number of secondary cases as well as individuals that result in matching the contract tracing.

The Big Data Architecture is quite important in ensuring the Sharing, data integration and analysis within the environment of big data. The Data Construction can result in an assessment of the real-time nursing of different public health surveillance systems. When information is collected from big data through social networks, it can easily be rebuilt in determining the outbreak of diseases (Sheng, Amankwah-Amoah, Khan, & Wang, 2020). Machine learning and Artificial Intelligence are considered the important expectations in screening the population of people that can help assess the infection risk. The temperature screening has generally been detecting the symptoms and the isolated persons expected to be infected. Moreover, it has been determined that the artificial intelligence-powered smartphone app will develop tracking the spread of COVID-19 in various Geographic areas.

5 Significant applications of big data in COVID-19 pandemic

Big data technology can help store a various quantity of information about the infected people with the virus of Corona. It generally aids in the understanding of the nature of these viruses. This technology is generally used to store the data of all types of cases pretentious by the COVID-19. This info

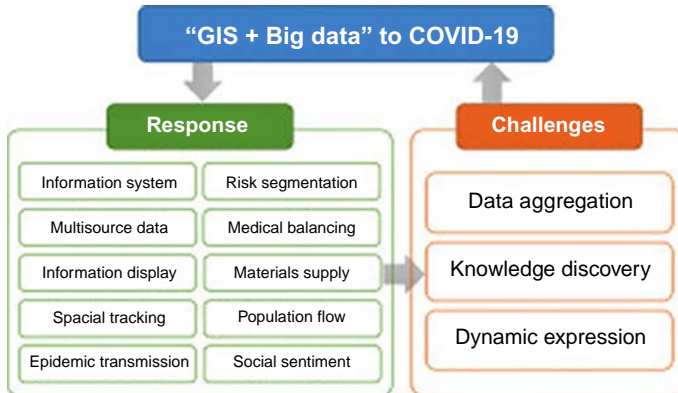


Fig. 7 Applications of big data in COVID-19. (Data from Zhou, C., Su, F., Pei, T., Zhang, A., Du, Y., Luo, B., et al. (2020). *COVID-19: Challenges to GIS with big data*. *Geography and Sustainability*, 1(1), 77–87.)

can review in identifying the cases and in helping and allocating the resources for the protection of the health of the public. See Fig. 7.

Several digital data modalities that include the location of the patient, patient-reported travel, proximity and current systems can generally be digitized and may be used in the generation of actionable insights at the demography and community level. Moreover, it has been determined that big data can help provide various amounts of evidence to health workers, epidemiologists, and scientists that can make informed decisions about fighting the virus. This data can mainly be used to track the virus on a continuous Global basis and help create innovation in the medical field. It can also help forecast the impact of COVID-19 in the whole population and particular area. Big data can also help provide opportunities and possible sources for the people that can help in having this stressful situation (Shorten, Khoshgoftaar, & Furht, 2021). This technology can help provide the data that can analyze disease transmission, monitoring of the health, movement, and prevention system. Letters of the big data will be acting as the medium for searching, tracking, and controlling the COVID-19 in the pandemic. It will take for the manufacturing and the development of the vaccine with having the knowledge. The model data helps to understand and offer the edge over the many other processes like corresponding homology models: product bi function and fold task service system for a specific target, duple also is downloaded from the protein data bank, find the symptoms, and predict the COVID-19 connected with it. Big data analysis can better understand

how the transmission could have happened and provide up-to-date updates and insights to both the stakeholders and the general public. With the help of data collection and analysis, together with lockdown, China suppressed the COVID-19 and implemented it with the help of artificial intelligence towards the spreading at a low rate.

The journal of Chinese medical association the article on the outbreak of COVID-19 was written by Wu, explain the first arrival of this virus in late 2019 December, the emergence of this virus is Wuhan, China, it out broke in some cities of China and then spread globally including the countries of Japan, United States, Philippines, Vietnam, India and many more. According to the ministry of health, infection is recognized as coronavirus disease or COVID-19 by WHO. This infection is a communicable disease under category 5. This disease is potentially zoonotic with a mortality rate to moderate. Its transmission may be spread through the droplets any kind of contact or if there is any kind regarding the control of stringent infection, or if there is no personal protective equipment. Presently there is no treatment for this virus all though the investigation is continuous (Tang, Schmitz, Persing, & Stratton, 2020).

With the continuous spreading of the infection worldwide, rapid tools of diagnosing, always handy repurposable drugs and instant containment measures Stopping the COVID-19 pandemic: a review on the advances of diagnosis, treatment, and control measures are of paramount importance. MD. Hasanul Banna Siam, in the journal of pathogens, explains that various kinds of vaccines, for example, RNA, DNA, and LENTIVIRAL, inactivated and viral vector is in trials in clinics. The very rapid deployable and innovative vaccine delivery systems have also been developing. So many countries have started going easy on the measures of lockdown, but the second wave of infection will need very strict as well as rational control policies so that the fatalities could be minimized (Siam, Banna, Nishat, Ahmed, & Hossain, 2020).

In the journal of clinical microbiology, the article on laboratory diagnosis of COVID-19: current issues and challenges elaborates that the outbreak of COVID-19 has had a very wide impact on clinical microbiology laboratories in the past several months. The current issues and challenges regarding the laboratory diagnosis of the infections caused by the very dangerous acute respiratory syndrome coronavirus 2 (SARS-COV-2) are widely covered. It has been analyzed that during the pre-analytical stage, the collection of the proper specimen of the respiratory tract at the right time and from the right anatomic site is essential for a fast and apt diagnosis of COVID-19.

In this article, a systematic review and meta-analysis have been done to investigate the optimum distance so that person-to-person transfer of the virus could be avoided and extensive uses of face masks and eye protection can largely reduce the transmission of this virus. This data has been searched through various data sources from the third of May 2020 inception with your language restriction for the studies and feasibility resource use and equity. The findings showed that there would be a lower transmission of this virus because physical distancing of at least 1 m or more will be maintained compared to a lower distance than that. The use of face masks resulted in a significant reduction in the risk of infection, the protection of the eye is also related to less infection (Chu et al., 2020).

Ahmad Elragal wrote on big data analytics in a literature review paper that huge amounts of data are available to decision-makers on hand in this era of information. Big data is referred to as data sets. These data sets are used to have very high amounts of velocity and variation, which makes them hard for handling using the same old-style tools and techniques. Because of such quick growth in data, solutions should be studied and should also be provided in audio for handling extracting value as well as knowledge from data-sets. Further, Big data analytics is a complicated process through which examination of big data is done to the revelation of information, for example, patterns that are not very clear, the trends regarding the market, the preferences of customers, which will be helpful to organizations for taking precise and apt decisions (Tsai, Lai, Chao, & Vasilakos, 2015).

With new age of big data which is impossible to be handled by the traditional data analytics due to huge quantities of data (Big data analytics: a survey by Chun-Wei Tsai, Chin-Feng Lai, Han-Chieh Chao, and Athanasius V. Vasilakos). The question now is how a great performance platform could be developed so that big data could be analyzed efficiently. An appropriate algorithm of mining could be designed to extract useful information from big data. On a wider scale, the technologies of data analytics and the techniques are helpful to organizations by providing a way in which analyzing data sets and gathering new and significant information could be done. The queries regarding intelligence can help answer the basic questions in relation to operations and the performance of the business (Elgendy & Elragal, 2014).

In this paper written by Bala M. Balachandran, the benefits and challenges included in the deployment of big data analytics are outlined. It is argued that the storage and computing necessities of big data analytics are supported by cloud computing. The combination of these very important

technologies will help in the enhancement process regarding big data mining improve the decision-making processes of the business. Organizations use big data analytics systems and software to make decisions that are data-driven so that there is improvement in the outcomes related to the business. The advantages include better marketing, greater opportunities for revenue generation, personalization of customers, and improved efficiency of operations. And these all benefits can be helpful in creating competitive advantages over the rivals (Balachandran & Prasad, 2017).

Machine learning is a method of artificial intelligence for discovering knowledge regarding making effective and intelligent decisions. Big data has huge effects on the value creation and discoveries related to science. Wang and Alexander (2016) introduce the methods in machine learning and the important technologies in big data and some kinds of applications inventor of machine learning. There has been a paradigm shift and the strategies of the firms globally, which is brought by social media. A resultant of it is the collection and message from so many varieties of social media platforms that have necessitated using this data for business intelligence purposes. This study has filled the gap in knowledge through the examination of the rules and implications regarding big data analytics business intelligence (Ram, Zhang, & Koronios, 2016).

The pandemic of COVID-19 has been labeled like the Dark swan event since it has affected a huge impact on almost all aspects of human life. The findings and the propositions in the study suggest that COVID-19 will become the substance of various long and short-term changes in policy and will also require researchers' hypothetical and pragmatic attention as published in the journal of business research (Verma & Gustafsson, 2020).

The noble coronavirus has spread to 210 countries worldwide which was declared a pandemic. It had a significant impact on the health systems, the economy, education, and social facets of present-day society. Since the transmission rate increased, many collaborative approaches in the stakeholders have developed innovative means of screening, detecting, and diagnosing coronavirus cases in humans. The importance of the computing models related to the 4th industrial rebellion technologies in achieving the wanted feat has been emphasized. However, there has been a wide gap in the two accuracies of the detection and the protection of the coronavirus cases. The tracing of the contacts of the person who is infected is also difficult. The following paper represents a review of the computing models

that can be adopted to enhance performance and predict global pandemic cases that are COVID-19. Tata official intelligence focuses more on big data, which is nature-inspired computing included in the present pandemic. The review has recommended that these non-natural intelligence models be used to detect the geysers of COVID-19. In the same way that forms of big data can also be applied for contact tracing. Nature-inspired computing models have illustrated amazing performances in the feature selection of the issues which are still open to exploration (Agbehadji et al., 2020).

Although the management scholars do distinguish the value of using the big data to comprehend, foresee and give response to upcoming events, there is a very inadequate overview of how so many different analytical techniques could be harnessed to help the guiding scholars in the researches and the global challenges raised in the pandemic related to COVID-19. Indeed the advancements, as well as the proliferation of so many technologies, have culminated. Various insights have been divided into the methods in the diagnostic, the predictor and the prescriptive analytics. These methods could be liberated to study “Black swan” events like the COVID-19 worldwide crisis and its various insinuations to the managers and policymakers aftermath (Sheng, Amankwah-Amoah, Khan, & Wang, 2020).

6 Research problem

The research problem of this project is dealing with data analysis and data mining during COVID-19 through machine learning in the addition of deep learning of procedure which is faced situations of COVID-19. In the current scenario of the epidemic, the people facing problems to analyze the existing solutions or strategies of COVID-19 impacts to minimize this impact can utilize new technologies like big data analysis with the help of machine education. Through the help of new technologies, data analysis and data mining of the pandemic is easy to identify and provides better outputs to solving problems of coronavirus disease.

7 Research questions

- What is the part of Machine Learning while dealing with the situation of COVID-19?
- How can a study of COVID-19 be leveraged by using Big Data Analytics and its method?

8 Objectives

The main objectives of this project are:

- **To explain the role of machine learning**

The role of machine learning techniques to tackle COVID19 crisis: A systematic review

They generally greatly alter the symptoms of a coronavirus that resulted in fluctuating from mild flu to severe respiratory suffering. The adoption of machine learning techniques has resulted in employing different skates that generally range from analyzing the disease spread. The various data types, including radiological images, drug databases and collection of the data from public health agencies, have generally been used for the prediction. It is only believed that machine learning can help predict all the three risks, such as infection risk, severity risk, and outcome risk. Machine learning founded CT analysis is also considered as the showing tool in the occurrence of COVID-19. It is required that the algorithms are required to be skilled in the full range of disease that includes asymptomatic. The occurrence of the COVID-19 has resulted in offering various exciting opportunities for the applied artificial intelligence research, as shown in [Fig. 8 \(Syeda et al., 2021\)](#).

- **To explain the various methods of data collection**

Entities can use big data analytics and similar software to make business decisions, which can help improve the outcomes related to those decisions. The advantages of them involve many new opportunities related to the generation of revenue, the personalization of customer experience, new and effective techniques of marketing, improvement in operational efficiency. With the use of an effective strategy, the benefits could be helpful in the creation of a competitive advantage over its rivals. The working of big data analytics could be explained in the following way: data analyst, the scientists of data, arithmeticians besides other professionals collect, then course, clean as well as examine the increasing capacities of the deal data which is structures and also many other forms of data which are not put into use by the analytics programs and the conventional business intelligence. The process of preparation of data is explained in the following four steps:

- Data collection is done by the professionals of data from various sources. Most times, it is like a mixture of data which is semi-structured and unstructured. Though every organization has different streams of data, some very common sources involve web server blogs, mobile

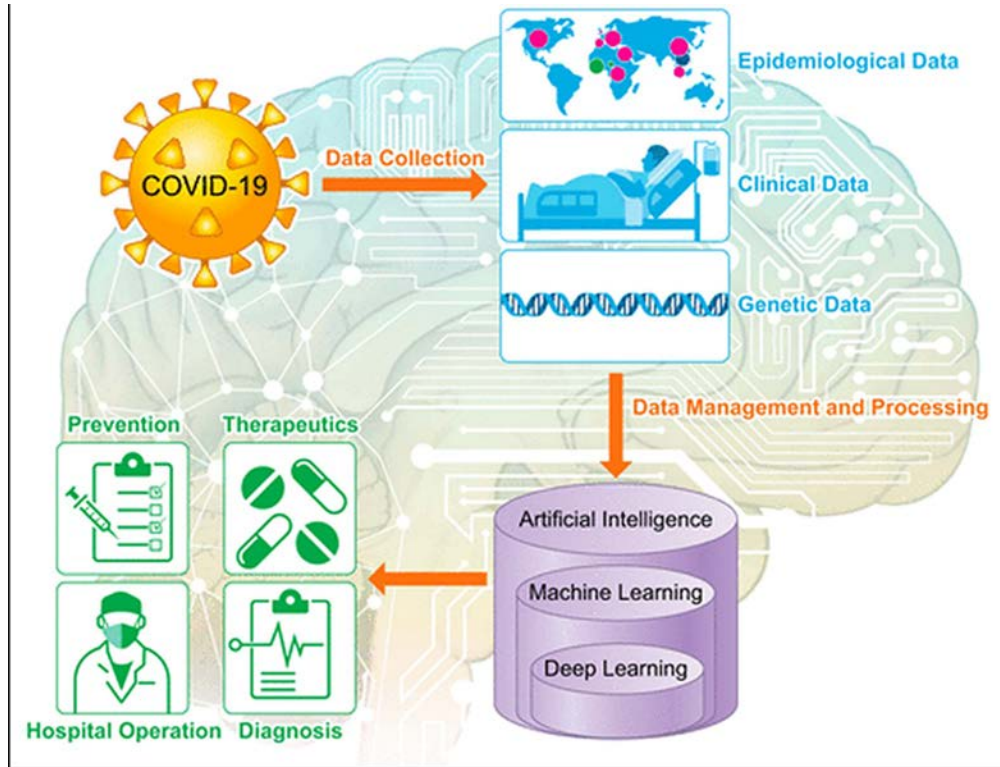


Fig. 8 Role of machine learning techniques to tackle COVID19 crisis. (Data from Alimadadi, A., Aryal, S., Manandhar, I., Munroe, P. B., Joe, B., & Cheng, X. (2020). Artificial intelligence and machine learning to fight COVID-19.)

applications, telephone records, the mechanism data taken by the devices linked to the internet of things, various cloud applications, and so on.

- Processing of data. Once the collection and storage of data are done in the data warehouse, its organization must be done by the professional of data. The partition, as well as the configuration of data, must be done properly for analytical queries. Proper processing of data contributes to higher performance from the analytical queries.
- In the third step, the cleaning of the data is done for better quality. Various tools of scripting or the software of the enterprise are used by the data professionals for scrubbing data. They try to find any errors or inconsistencies present, such as the duplication of data, any mistakes in formatting, organization or tidying up the data.

- **To explain different algorithms utilize in machine learning**

Some very popular methods of machine learning are as follows:

- Supervised learning: the uses of labeled examples train these algorithms. For example, for an input name, the output that is desired is already known—an example of an equipment piece that might have points of data labeled. Moreover, F means unsuccessful and R means runs. The learning algorithm receives a set of inputs along with the correct corresponding output, and learning in this algorithm is done by creating a comparison of the actual output with the correct output for finding errors. The modification of the model is then done accordingly. By using approaches like sorting, forecasting, and gradient boosting unsupervised wisdom, the designs are protected by the values regarding the label on the unlabeled data, which is extra. Supervised learning is most commonly used in applications where the prediction of historical data is likely to predict upcoming events. For instance, anticipation could be done when dealings in the credit card could be fraudulent or when the customer of insurance has chances of filing a right. See [Fig. 9](#).
- Unsupervised learning: it is cast-off as in contradiction of the data with no historical labels. No “right answer” is expressed by the system. It is up to the algorithm regarding what is to be shown. The main goal is the exploration of the data and searching for some kind of structure within. This kind of unsupervised learning is fine suited to transactional kinds of data. Given examples, identifying a segment of customers with very akin attributes can be treated in a similar way when it comes to marketing campaigns. See [Fig. 10](#).

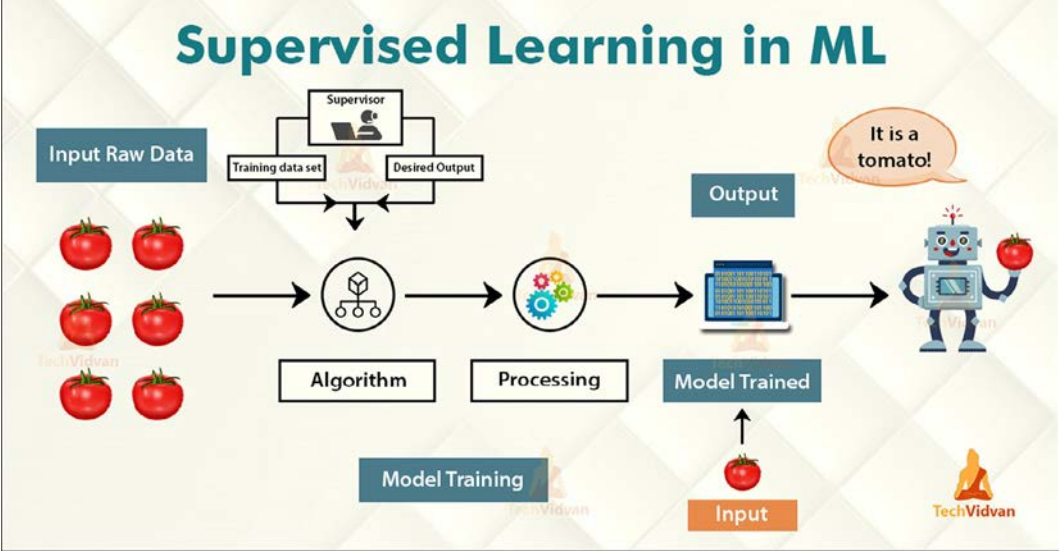


Fig. 9 Supervised learning. (<https://www.linkedin.com/pulse/brief-review-artificial-intelligence-its-subfields-dwaipayana-saha/>)

- Reinforcement learning: this is furthestmost used for gaming, steering as well as robotics. Through the methods of trials and errors, this algorithm can discover what kinds of actions will help yield the biggest recompenses. There are three main components regarding this kind of learning which are: the agent (the choice maker), the atmosphere (everything is interacted by the agent) and the actions (everything in which the agent can do). The main purpose for the agent is to decide which actions will help maximize the reward, which was expected over a specific period. See Fig. 11.
- Semi-supervised learning: its use is similar to that of supervised learning, but the only difference is that both labeled and unlabeled data are used for the training, mostly some quantity of data labeled in combination with a huge amount of data quantity of data which is unlabeled. The usage of this learning is in various methods like classification, prediction as well as regression. When the cost related to the labeling is very high such that a fully labeled process of training could be allowed, in such situations, semi-supervised learning is very effective and useful. See Fig. 12.

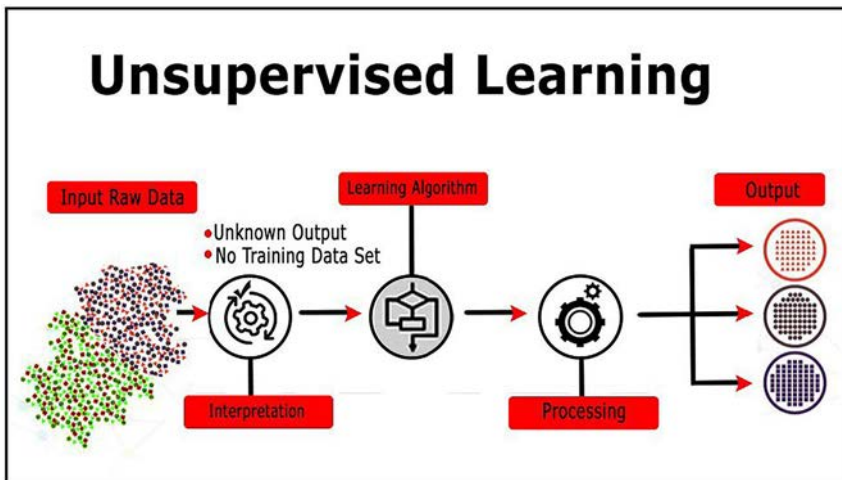


Fig. 10 Unsupervised learning. (<https://towardsdatascience.com/into-to-pca-k-means-clustering-unsupervised-learning-algorithms-5cc5acea274d>.)

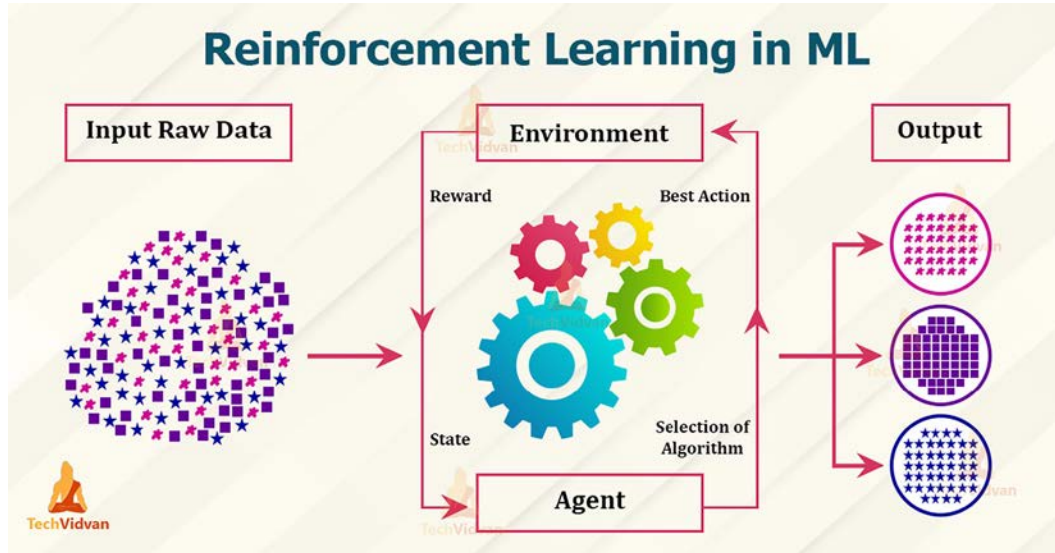


Fig. 11 Reinforcement learning in ML. ([https://techvidvan.com/tutorials/reinforcement-learning/.](https://techvidvan.com/tutorials/reinforcement-learning/))

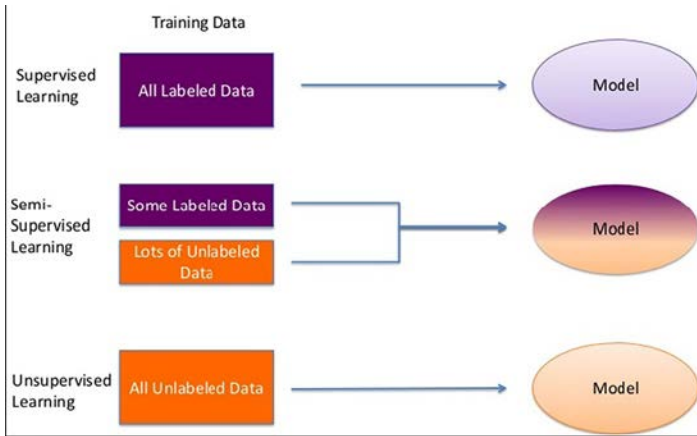


Fig. 12 Semi-supervised learning. (<https://www.analyticsvidhya.com/blog/2017/09/pseudo-labelling-semi-supervised-learning-technique/>.)

- **To explain the utilization of BDA and machine learning in the situation of COVID-19**

According to the author (Jia et al., 2020) design and development of new generation architecture and technology can be helpful for BDA or Big Data Analytics to get maximum velocity and examine and discover various variations and high volume of data for extracting values from them. The increasing popularity of information technologies such as big data technology, social networks, cloud computing or IoT, or the Internet of Things has been used widely in healthcare, supply chains, logistics, and other industries. BDA is helpful and convenient in decreasing fear by revealing the epidemic situation in real-time and clarifying rumors regarding the COVID-19. Model of BDA or analysis of big data can be used to estimate the impact of this situation on social development, economic, political development and assist the government in making suitable changes.

- **To explain domain adaptation and big data**

As per the words (Jia et al., 2020) of three mechanisms that can be used for controlling and preventing COVID-19 situation by big data application are prevention of situation, response to situation and recovery from the situation. With the use of visual analysis technology, the relationships between massive datasets can be uncovered. It enables the investigators to attain visual cognition that is more intuitive and sufficient support in the decision-making process. Visualization is achieved with the help of GIS or Geographical Information Systems. The government uses sources of bid data for conducting pictorial analysis of monitoring the widespread

situation, hospital enterprises, medical resources, and screening for making decisions. Government can improve mechanisms used by them for an epidemic response that are based on analytics regarding big data. Big data technology is used to identify disasters, supporting the decision, communication and coordination, and technical support. For identification of disaster, predictive scrutiny of dynamic models of infectious disease is combined with data to predict how critical a situation is and provide support. Systems of geographic situation and examination of Graph database are used for tracking persons who are infected and who are their contacts. Government can establish a mechanism for repairing an epidemic based on analysis of big data. Big data is used for eliminating fear, audit assessment, recovery and adjustment of policy.

9 Methodology

In this research, Experimental Research methodology is used to find the best algorithm. An experiment is a kind of research method in which there will be manipulation of one or more independent variables and missing their impact on one or more dependent variables. An experimental design is a process of the development of sets of procedures for testing the hypothesis. Experimental research methodology is being conducted with the help of a scientific approach by utilizing two sets of variables. The first set will play a role of a constant that has been used for measuring the differences of the second set. This research methodology will help establish a noticeable effect and cause. This type of research methodology will help collect information from present and potential customers by using the sampling method. With the help of various kinds of surveys, questionnaires and the outcomes of those will be depicted in the form of numbers. See [Fig. 13](#).

The data collection methods used for current research involve no data collection and the algorithms discussed have been sourced from various existing literature.

9.1 Techniques

The data and other relevant information are gathered through secondary search. It is a kind of research method which includes the utilization of already present data. The data present has been summarized and collected to enhance the overall effectiveness of the research. This search involved research material that is being published in a research report and similar kinds



Fig. 13 Experimental design. (<https://dewezanake.vercel.app/index/view-17-one-shot-case-study-research-design-example/>.)

of documents. It is more cost-effective than primary research as it makes utilization of already present data.

10 Algorithm

The algorithm that has been used in this research is Naive Bayes. It is a kind of simple and supervised machine learning algorithm that utilizes the Bayes theorem that involves stronger and independent assumptions among the features for procuring results. This algorithm assumes that every input variable is an independent variable. It is a kind of naive assumption for making results about real-world data. See [Fig. 14](#).

This algorithm is a kind of classification algorithm based on the Bayes theorem. It is not a sole algorithm but involves the family of algorithms where all algorithms share a common principle: each pair of features is classified independently. Bayes theorem states that the probability of occurrence of an event is given by the probability of another event that exists already. See [Figs. 15 and 16](#) for formulas.

Naïve Bayes is generally a classification algorithm based on the Naïve Bayesian Algorithm and strong Independence assumptions. This type of algorithm simplifies learning by undertaking features that are independent of provided classes. Naive Bayes is a type of Bayesian decision theory subset. The name naive demonstrates the formulation undertakes some naïve assumptions. Naïve Bayes is among the simpler probabilistic classifier. This displays amazingly well in several Real-world applications. The Naïve Bayes classifier communicates to the Bayesian network that can be demonstrated with the below equation.

Classification is generally a fundamental issue within data mining and machine learning. In classification, the major learning algorithm goal is to develop a classifier given a training set of examples with class labels. Such algorithms are utilized in diverse real-life scenarios such as spam filtration, sentiment analysis, text classification, and recommendation system. The advantage of the Naïve Bayes algorithm is that it is a relatively easier algorithm to build and understand. It is also faster to predict classes compared to other classification algorithms, and it can be easily and with the use of a small data set. In general terms, this classifier assumes that the existence of a certain feature in class concerns the existence of another feature. The Naïve Bayes model is easy to use and build for large datasets and its simplicity. The Naïve Bayes is called for outperforming the highly sophisticated classification methods. During the Independence assumption holds,

Naive Bayes

In machine learning, naive Bayes classifiers are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong (naive) independence assumptions between the features.

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

using Bayesian probability terminology, the above equation can be written as

$$\text{Posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

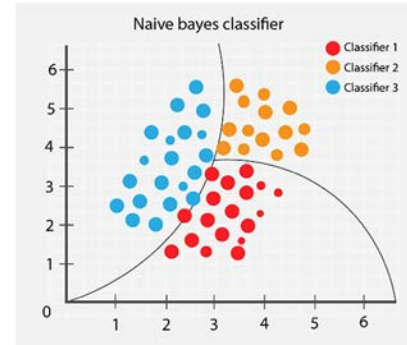


Fig. 14 Naive Bayes. (<https://coderzcolumn.com/tutorials/machine-learning/scikit-learn-sklearn-naive-bayes>.)

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Probability of B occurring given evidence A has already occurred Probability of A occurring
 Probability of A occurring given evidence B has already occurred Probability of B occurring

Fig. 15 The formula used.

$$Pr(c, x_1, \dots, x_m) = Pr(c) \cdot \prod_{i=1}^m Pr(x_i | c)$$

Fig. 16 Naïve Bayes.

this classifier performs a suitable comparison to the other models such as Logistic regression and one further requires fewer data. This algorithm performs well in the situation of categorical input variables compared to the other type of numerical variables. Other applications of such Naïve Bayes algorithm include multiclass prediction and real-time prediction. Moreover, the Naïve Bayes algorithm is not always consistent and dependent on the quality of good datasets as the limitation. Please see Fig. 17 for the input and output of Naïve Bayes analysis and Fig. 18 for showing different results and analyses for COVID-19.

11 Conclusion

11.1 Big data

Big data signals an emerging emphasis on the rights of the data-centric approach to research that has to integrate, mobilize, visualize, and

Formula used

```
In [3]: from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
X[:,0] = le.fit_transform(X[:,0])

In [4]: from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.20, random_state = 0)

In [5]: from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

In [6]: from sklearn.naive_bayes import GaussianNB
classifier = GaussianNB()
classifier.fit(X_train, y_train)

Out[6]: GaussianNB()

In [7]: y_pred = classifier.predict(X_test)

In [8]: y_pred
Out[8]: array([0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1,
               0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
               1, 0, 0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1,
               0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1], dtype=int64)

In [9]: y_test
Out[9]: array([0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1,
               0, 1, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
               1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 1,
               0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 1], dtype=int64)
```

(A)

Fig. 17 cont'd

Naïve Bias Classification Input

```
In [3]: from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
X[:,0] = le.fit_transform(X[:,0])

In [4]: from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.20, random_state = 0)

In [5]: from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

In [6]: from sklearn.naive_bayes import GaussianNB
classifier = GaussianNB()
classifier.fit(X_train, y_train)

Out[6]: GaussianNB()

In [7]: y_pred = classifier.predict(X_test)

In [8]: y_pred

Out[8]: array([0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1,
0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
1, 0, 0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1,
0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1], dtype=int64)

In [9]: y_test

Out[9]: array([0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1,
0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 1,
0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 1], dtype=int64)
```

Output

```
In [12]: from sklearn.metrics import confusion_matrix, accuracy_score
print("The Confusion Matrix : ")
print(confusion_matrix(y_test, y_pred))
print("The Accuracy Score is : ", accuracy_score(y_test, y_pred))

The Confusion Matrix :
[[56  2]
 [ 4 18]]
The Accuracy Score is : 0.925
```

```
In [ ]:
```

Analysis

(C)

Fig. 17 Inputs and command-line outputs of running Naïve Bias (A and B).

Visualization of COVID-19 Dataset:

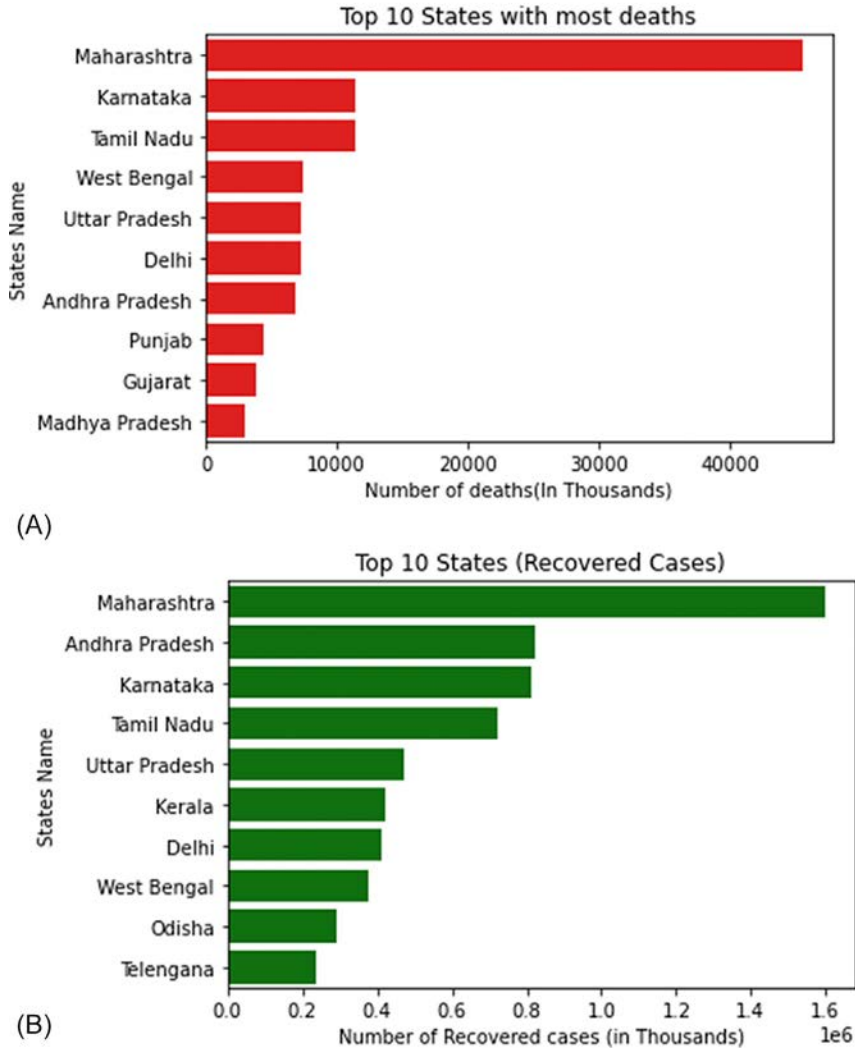
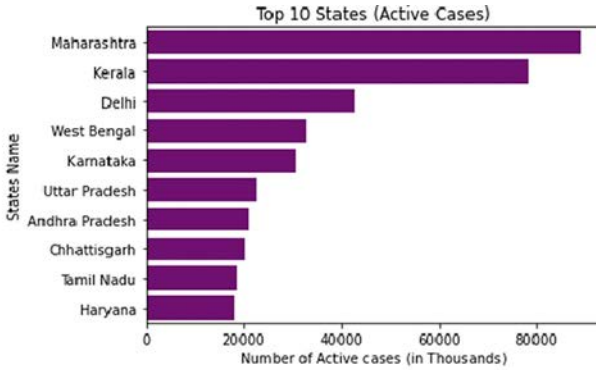


Fig. 18 Cont'd



(C)

State	Recovered	Deaths	Confirmed	Active
Maharashtra	1597255	45560	1731833	89018
Karnataka	811581	11453	853796	30762
Andhra Pradesh	820234	6828	847977	20915
Tamil Nadu	720339	11415	750409	18655
Uttar Pradesh	473316	7281	503159	22562
Kerala	422410	1771	502719	78538
Delhi	410118	7228	459975	42629
West Bengal	376696	7452	416984	32836
Odisha	292477	1469	305000	11054
Telengana	235950	1393	254666	17323
Bihar	216940	1162	223917	5815
Rajasthan	198139	2019	217151	16993
Assam	203308	954	209633	5371
Chhattisgarh	183190	2507	205923	20226
Haryana	170250	1960	190323	18113

(D)

Fig. 18 Cont'd

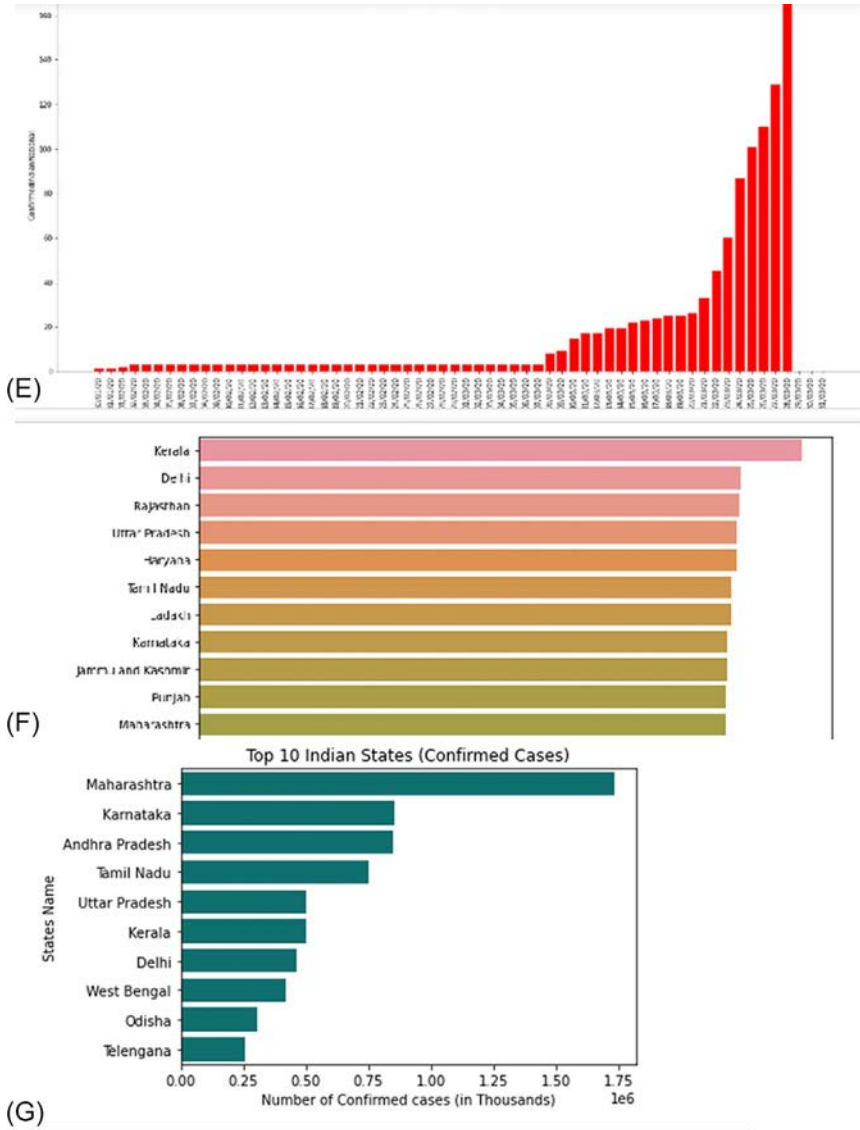


Fig. 18 Cont'd

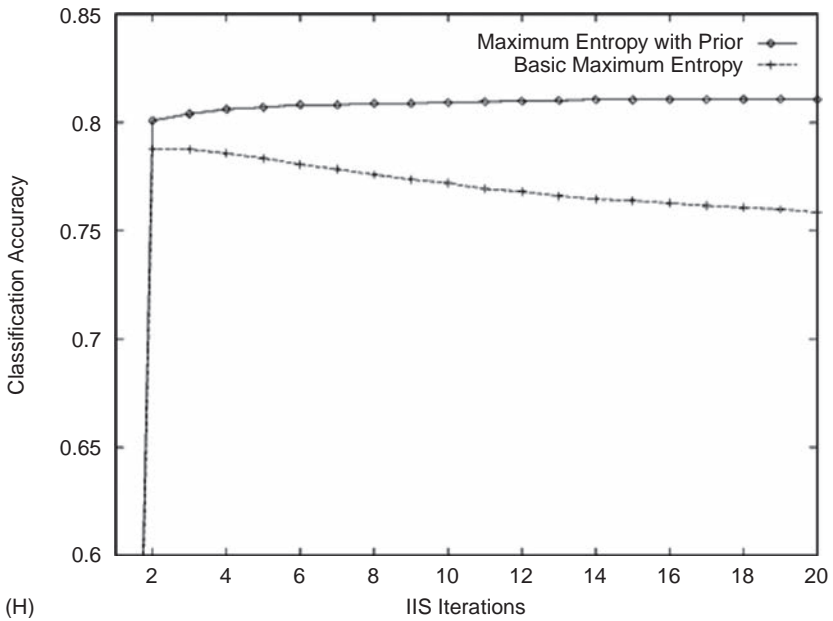


Fig. 18 Different results and analyses for COVID-19 (A–H).

disseminate the data that is viewed as contributing to the discovery. The exposure of the data centrism highlights the challenges involved in classifying, gathering and data interpretation, technologies, and the concept surrounding this process—tools such as applications and measurement instruments that produce a huge amount of data in the digital format. Data are available immediately for mentioning through the online platforms available or accessible to anybody with a broadband connection in a matter of seconds. The data access is carried through technical, ethical conceptual and legal implications and even if the access is available. The computational and mathematical tools developed to determine the big data or an opaque in their assumption and functioning lead to credibility and the scientific meaning that may difficult to access. This resulted in increasing the very related to the big data science that may be grounded upon and result in supporting the process of human making ingenuity Hostage and ultimately unintelligible intelligence.

Moreover, it has been determined that the most confronting expectation in relation to big data science has been discussed in this article: the extent to which it results in a deviation from an understanding of rationality based on

individual agency and abilities. The power of particular data set results in ending the knowledge that generally lies in the extent to which it is linked to others. It results in high epistemic value to digital objects such as sequencing data on GPS location and what generally makes aggregation of data from a wide variety of sources into an effective surveillance tool. Dissemination channels and data products such as government databases, mission media and Research repositories generally read in the Interlinked and globalization distributed network that tends to function through skills and expertise. The distributed nature of making decisions is mainly involved in developing big data and analytics. It results in it impossible for any particular individual to retain oversight over the scientific significance, quality, and potential social impact of the knowledge produced.

The analysis of the big data main constitutes the ultimate instance of the cognitive system. Many groups' individuals and institutions result in ending the responsibility for the concept interpretations and outcomes of specific data use. The main key challenge related to big data government is finding the mechanism for allocating the responsibility under the acrostic from its network so that unwarranted and erroneous decisions can be taken or sanctioned. Thinking about the particular history use of the term processing can generally increase the Philosophers in avoiding historical, contextualized approaches concerning the question of evidence.

11.2 Machine learning

View detailed analysis of machine learning, and it has been analyzed that it opens up a parallel opportunity for the particular organization and enables efficiency, automation and innovation. It has been analyzed that the five pillars of Steve architected Framework based on machine learning can provide best practices in architecture for operating and building secure, reliable, efficient and cost-effective machine learning workloads in the Clouds.

Such operation automation requires teams of cross functions and areas of expertise to maintain and the major functions have full automation capabilities. For reliability, there is a robust response when there is a failure of the system by taking advantage of self-healing based on automatic scaling and keeping track of all kinds of artifacts by changing the functional system that can be automatically rebuilt.

Workloads of machine learning should be physically secured by using authorization and authentication tightly controlled with user identities, allowing users to access data and artifacts. The application will help in

protecting the sensitive information of the organization and compliance requirements at every layer. Machine learning and landscape generally involve the ecosystem of tooling and growing processes and maturing.

After the coronavirus outbreak, the medical industries and scientists around the globe have worked in fighting against this pandemic and searching different production process methods and Rambha screening forecasting contact tracing and growth of the vaccine or drugs with the more reliable and accurate process. That official Intelligence and machine learning is considered a promising method that helps in improving various Health Care providers. This research has addressed Singh audition studies that generally apply different technologies to supplement the researchers in different angles and speak of challenges and trouble using different algorithms to assist the medical experts in related problems. This paper also has discussed the suggestion about conveying research on artificial intelligence-based model design policymakers and medicinal experts on this is a few and those met in the present situation while undertaking the coronavirus. Through this, it has been analyzed the use of modern technology such as artificial intelligence, and Artificial Intelligence can help refine the predictions training content using drugs and development and forecasting through extreme reliability? The power of particular data set results in ending the knowledge that generally lies in the extent to which it is linked to others. It results in high epistemic value to digital objects such as sequencing data on GPS location and what generally makes aggregation of data from a wide variety of sources into an effective surveillance tool. The application of technology in the research results in improving various algorithms and is more from this potential and advanced among the different learning algorithms. However, the current agency generally requires that there should be an improved model and the higher and performance that result in protecting and screening the coronavirus to use a different kind of diseases and analyzing its clinical and demographic info of the infected patients. Finally, it has been determined that machine learning and artificial intelligence are essential in improving the medication treatment screening and prediction and reducing the intervention of humans in the medical practice.

11.3 COVID-19

It is difficult to predict the coronavirus spreading across the world. The humanitarian health and the policies of the social economy will be adopted by different countries, which will analyze their strength and speed of

recovery. Different Rock present in brief gives the country guidance as they are progressing through the different phases of crisis and also to the international community. There must be a response that is human-centered and grounded in solidarity. The International Labour values and contained the guidelines for safeguarding the decent work that is even valid in the context of the catastrophe of a pandemic. The main emphasis is on responding to the crisis that needs to ensure that human rights should be respected. And the rule of law should include the respect for principles that are basically at exertion for global labor standards. The standard guidelines for employees at work and safety may undergo public security, non-discrimination preparations of emergencies and isolation of infected workers and workers under quarantine are required to the rapid response to ease in stouter recovery from the crisis. It is required that there should be coordinated Global effort that can help in supporting the countries that can lead to and is not sufficient in fiscal space and financing the social policy in particular social protection system. It has been determined that without proper long-term structure changes, the inequalities exposed in deep-rooted crises will intensify. This will help take the immediate effect of prices and has unique opportunity to adopt a policy that thinks that achieving human-centered in Social Justice Future of work. The virus is transmitted between the people to the other person. It shows the symptoms and 2–14 days depending upon the incubation period. As per the symptoms and this sign of middle to moderate, moderate are for fatigue, fever and dry cough, and shortness of breath. Tiredness and fever also occur. Individuals with other diseases like diabetes, heart disease, and asthma are more deadly viruses and may become ill. While covering and sneezing the mouth in noise with the help of the tissue that is disposable and avoiding contact with the year mouth and nose can help in the prevention. The world has shock due to COVID-19 because of its non-availability of drugs or vaccines. Different researchers are working on the deadly virus to make do vaccines for it. The clinical report has been named COVID, ARDS. The advancement of technology generally has a rapid effect on every field of life, whether medical or other—tools such as applications and measurement instruments, generating a huge amount of data in the digital format. Data are available immediately for mentioning through the online platforms, which will be available or accessible to anybody with a broadband connection in a matter of seconds. It has been determined that artificial intelligence is considered a promising result in the Healthcare sector as it helps analyze decision-making and data. It has been determined that COVID-19 has generally affected more than a hundred

countries. Moreover, it has been determined that people worldwide are vulnerable to different consequences in the future. It is quite important in developing the control system that will result in the detection of the coronavirus. One of the main solutions and controlling the current problem is diagnosing the disease that can be easily done through artificial intelligence tools. The occurrence of the COVID-19 has generally shocked the world as there is no availability of drugs or vaccines. Moreover, it has been determined that various researchers are mainly working to win over COVID-19 viruses. Various features like IDF and TF are extracted from the clinical reports. The machine learning algorithm is generally used in the classification of clinical reports in two different classes.

Acknowledgment

This work is partly supported by VC Research (VCR 0000155) for Prof Chang. The second co-author has obtained permissions for use of images before the completion of this research on the behalf of the team.

References

- Agbehadji, I. E., Awuzie, B. O., Ngowi, A. B., & Millham, R. C. (2020). Review of big data analytics, artificial intelligence and nature-inspired computing models towards accurate detection of COVID-19 pandemic cases and contact tracing. *International Journal of Environmental Research and Public Health*, 17(15), 5330.
- Balachandran, B. M., & Prasad, S. (2017). Challenges and benefits of deploying big data analytics in the cloud for business intelligence. *Procedia Computer Science*, 112, 1112–1122.
- Chu, D. K., Akl, E. A., Duda, S., Solo, K., Yaacoub, S., Schünemann, H. J., et al. (2020). Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: A systematic review and meta-analysis. *The Lancet*, 395(10242), 1973–1987.
- Corsi, A., de Souza, F. F., Pagani, R. N., & Kovaleski, J. L. (2020). Big data analytics as a tool for fighting pandemics: A systematic review of the literature. *Journal of Ambient Intelligence and Humanized Computing*, 1–18. <https://doi.org/10.1007/s12652-020-02617-4>.
- Elgendy, N., & Elragal, A. (2014, July). Big data analytics: A literature review paper. In *Industrial conference on data mining* (pp. 214–227). Cham: Springer.
- Haleem, A., Javaid, M., Khan, I. H., & Vaishya, R. (2020). Significant applications of big data in COVID-19 pandemic. *Indian Journal of Orthopaedics*, 54, 526–528.
- Jia, Q., Guo, Y., Wang, G., & Barnes, S. J. (2020). Big data analytics in the fight against major public health incidents (including COVID-19): A conceptual framework. *International Journal of Environmental Research and Public Health*, 17(17), 6161.
- Lalmuanawma, S., Hussain, J., & Chhakhuak, L. (2020). Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review. *Chaos, Solitons & Fractals*, 139, 110059.
- Ram, J., Zhang, C., & Koronios, A. (2016). The implications of big data analytics on business intelligence: A qualitative study in China. *Procedia Computer Science*, 87, 221–226.

- Sheng, J., Amankwah-Amoah, J., Khan, Z., & Wang, X. (2020). COVID-19 pandemic in the new era of big data analytics: Methodological innovations and future research directions. *British Journal of Management*. <https://doi.org/10.1111/1467-8551.12441>.
- Shorten, C., Khoshgoftaar, T. M., & Furht, B. (2021). Deep learning applications for COVID-19. *Journal of Big Data*, 8(1), 1–54.
- Siam, M., Banna, H., Nishat, N. H., Ahmed, A., & Hossain, M. S. (2020). Stopping the COVID-19 pandemic: A review on the advances of diagnosis, treatment, and control measures. *Journal of Pathogens*, 2020, 1–12.
- Syeda, H. B., Syed, M., Sexton, K. W., Syed, S., Begum, S., Syed, F., et al. (2021). Role of machine learning techniques to tackle the COVID-19 crisis: Systematic review. *JMIR Medical Informatics*, 9(1), e23811.
- Tang, Y. W., Schmitz, J. E., Persing, D. H., & Stratton, C. W. (2020). Laboratory diagnosis of COVID-19: Current issues and challenges. *Journal of Clinical Microbiology*, 58(6), e00512–20.
- Tsai, C. W., Lai, C. F., Chao, H. C., & Vasilakos, A. V. (2015). Big data analytics: A survey. *Journal of Big Data*, 2(1), 1–32.
- Verma, S., & Gustafsson, A. (2020). Investigating the emerging COVID-19 research trends in the field of business and management: A bibliometric analysis approach. *Journal of Business Research*, 118, 253–261.
- Wang, L., & Alexander, C. A. (2016). Machine learning in big data. *International Journal of Mathematical, Engineering and Management Sciences*, 1(2), 52–61.