IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

USING BIG DATA ANALYTICS TO CONTROL COVID-19 PANDEMIC

P.V.Naga Lakshmi,

Assistant Professor, Department of Master of Computer Applications,
Loyola Academy, Old Alwal, Secunderabad, Telangana, India.

Abstract: **The COVID-19 epidemic has caused an outsized number of human losses and havoc the economic, societal, social, and health systems around the world. Controlling such epidemic requires understanding its behavior and characteristics, which may be identified by collecting and analyzing the related big data. However, because of the vast amount of data available on COVID-19 from various sources, there is a requirement to review the roles of massive data analysis in controlling the spread of COVID-19, presenting the most challenges and directions of COVID-19 data analysis. Therefore, during this paper, we conduct a literature review to highlight the contributions of several studies within the domain of COVID-19-based big data analysis.**
The study presents as taxonomy of several applications accustomed manage and control the pandemic.
Moreover, this study discusses several challenges encountered during analyzing COVID-19 data.
The findings of this paper suggest valuable future directions to be considered for further research and applications.

Index Terms - **big data; big data analytics; AI (AI); 2019 novel corona virus disease (COVID-19); healthcare**

I. INTRODUCTION

The novel corona virus Covid-19 was originated in China in early December 2019 and has rapidly spread too many countries round the globe, with the amount of confirmed cases increasing on a daily basis. Covid-19 is officially an endemic. It's a unique infection with serious clinical manifestations, including death, and it's reached a minimum of 124 countries and territories. On 30 January 2020, the globe Health Organization (WHO) declared the spread of the COVID-19 pandemic as a reason for concern and required raising the amount of health emergencies. the primary case of COVID-19 in India, which originated from China, was reported on 30 January 2020.[1] India currently has the most important number of confirmed cases in Asia.[2] May 2021, India has reported the second-highest number of confirmed cases in the world (after the US) with 25.4 million COVID-19 infection reported cases and 283,248 deaths on 19 May 2021.[3][4] Lockdowns were announced first in Kerala on 23 March 2021, and rest of the country on 25 March 2021. By mid-May 2020, 5 cities got half of the reported cases in the country: Mumbai, Delhi, Ahmedabad, Chennai and Thane.[5] On 10 June, India's Covid-19 patients recoveries exceeded active cases for the primary time.[6] Infection rates began to call September, together with the amount of recent and active cases.[7] Daily cases peaked mid-September with over 90,000 cases reported per-day, dropping to below 15,000 in January 2021.[8].

A second wave beginning in March 2021 was much larger than the primary, with shortages of vaccines, hospital beds, oxygen cylinders and other medicines in parts of the country.[9] By late April, India led the globe in new and active cases. On 30 April 2021, it became the primary country to report over 400,000 new cases during a 24-hour period.[10][11] Health experts believe that India's figures are underreported because of several factors.

A regular monitoring and remote detection system for people will assist within the fast-tracking of suspected COVID-19 cases. Moreover, using such systems will generate a huge amount of knowledge, which can provide many opportunities for applying big data analytics tools [12] that are likely to enhance the extent of healthcare services. There is an outsized number of open-source software like the large data components for the Apache project [13], which are designed to work during a cloud computing and distributed environment to assist within the development of massive data-based solutions. Furthermore, there are several key
characteristics of massive data called the Six V's [14], namely, Value, Volume, Velocity, Variety, Veracity, and Variability. However, the first definition of the large data key characteristics considers only three Vs, namely Volume, Velocity, and Variety [15].

All big data characteristics apply to the data acquired from the healthcare sector, which increases the tendency to use big data analysis tools to enhance sector services and performance.

There are wide applications of massive data analytics within the healthcare sector, including genomics [16], drug discovery and clinical research [17], personalized healthcare [18], gynecology [19], nephrology [20] and a number of other applications
found within the literature. However, during this paper, we present the contributions of the foremost important review papers found within the literature that cover the sector of massive data in healthcare.

Promising wearable technology is predicted to be one among the first sources of health information, given its large widespread availability to people and acceptance by people. Supported a survey conducted in January 2020, 88% of 4600 subjects included within the study indicated a willingness to use wearable technology to live and track their vital signs. While 47% of chronically ill patients and 37% of non-chronically ill patients reported willingness to frantically distributed patient's health information with healthcare research organizations. Of an equivalent group, 59% said they might likely use AI (AI) -based services to diagnose their health symptoms [21]. People sharing such data routinely will greatly increase the volume of knowledge, which involves getting to design and implement data analysis tools and models during this sector.

Several studies used big data for sentiment analysis, like Reference [22], which linked between social media behavior and politics, opinions, and expressions. The study consisted of a large representative survey conducted on 62.5% of adults from Chile and it showed the large effect of social media on changing people's opinions regarding political views and elections. Similarly, the authors of Reference [23] had studied how the management responding to customer satisfaction online review affects the selection of the purchasers for some facilities or hotels. It showed a direct correlation between the response and customer satisfaction. The authors of Reference [24] had analyzed the classification techniques, including deep and convolution techniques.

They discussed several challenges in identification associated with language characteristics, scripts, and therefore the lack of datasets. Also, the authors of Reference [25] had reviewed and analyzed the latest developments, capabilities, and profits papers about big data analytics. Their survey showed that big data can aid business oriented industries in many areas including prediction, planning, managing, decision-making, and traceability.

The limitation of their study is that the data sources, which were hard to seek out thanks to privacy and conservation of the knowledge .The authors of Reference [26] had analyzed various papers about mathematical models to extend the efficiency in detecting and predicting COVID-19. Their survey suggested using AI to detect COVID-19 cases, big data to trace cases, and nature-inspired computing (NIC) to pick suitable features to extend the accuracy of detection. Some surveys studied heart-related diseases and d some recommendations and guidelines, like Reference [27], help people in understanding coronary failure causes, symptoms, and therefore the most affected group. They declared that coronary failure can escalate the patient's injuries, especially the ones with serious illnesses.

Examining health data in real-time with the work of AI techniques will have a important role in predictive and preventive healthcare system. For example, it'll help to predict the sites of infection and therefore the flow of the virus. it'll also help in estimating the requirements of beds, healthcare specialists, and medical resources during Covid pandemic also as well in the diagnosis and characterization of the corona virus [28].Many reviews within the literature have analyzed big data analytics in healthcare system from various aspects. During this paper, we specialize in identifying the applications of massive data analytics for COVID-19.

2. Applications of Big Data Analytics in COVID-19

The spread of the COVID-19 global pandemic has generated a large and different amount of data, which is increasing rapidly day to day. The rapid spread of COVID-19 meant that hospitals had to organize for the worst. During a system that's already strained, the potential for waves of highly contagious patients can only translate to disaster.

With big data analytics tools, organizations were ready to track and monitor the utilization of critical resources. Definitive Healthcare, in partnership with Esri, launched an interactive data platform allowing people to research US single bed capacity, also as potential geographic areas of risk, during the COVID-19 outbreak.

The platform shows the situation and number of licensed beds, staffed beds, ICU beds, and total bed utilization within the US. The rapid spread of COVID-19 has brought together doctors, researchers and data scientists to seek out an answer. Scientists are using sophisticated technologies like big data analytics, machine learning, and tongue processing for tracking the virus and learning more about it.

As many patient data is being stored now, it's difficult to research each record and determine an answer to curb the virus. This is where big data comes in. Big data is becoming a robust tool in analyzing these datasets and identifying patterns which can help in COVID-19 detection and recovery. Do we know how Big Data is helping in the fight against COVID-19?

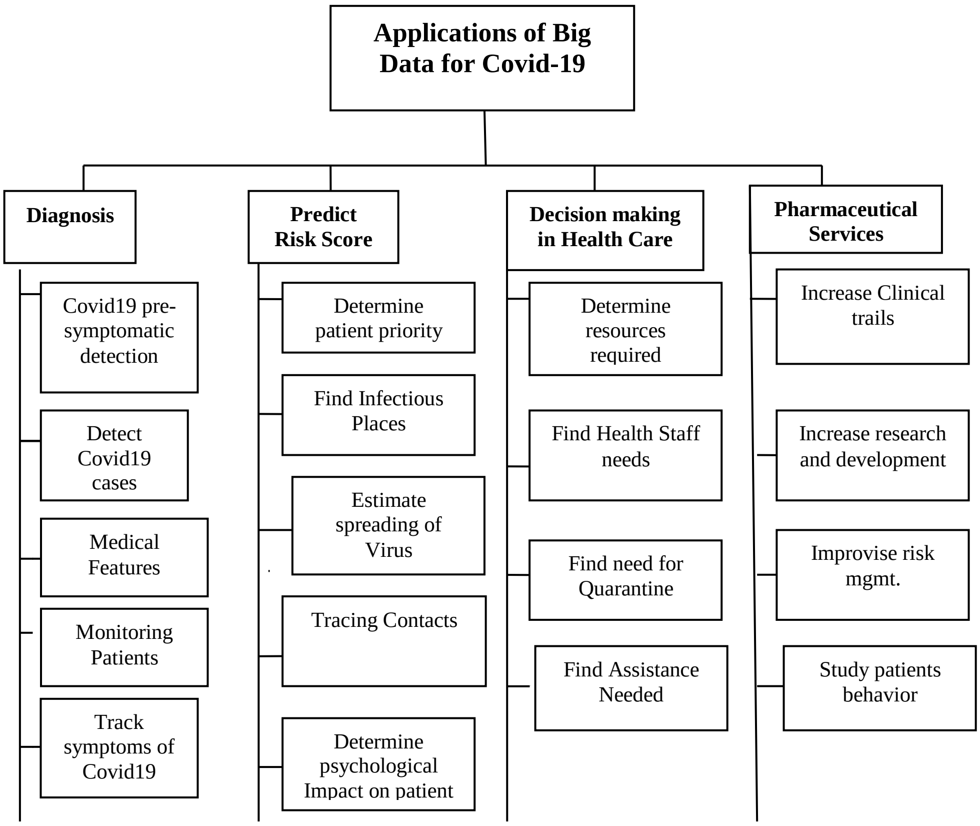
The big data plays important role in COVID-19 starts from the initial step i.e. detection. BluDot Toronto-based big data start-up detected some unusual pneumonia cases in Wuhan, China in December 2019. They did it by using their big data algorithm that pulled data from a spread of sources. The algorithm analyzed data from health records, airline ticketing data, government notices, news reports and disease networks to predict the increase of an illness.

Analyzing the spread of COVID-19 using airline ticketing data, BluDot was ready to predict the spread of COVID-19 from Wuhan to other Asian cities. Apart from this, mobile data is additionally used for tracking where the virus might spread. Location statistics are also useful. For example, big data tools can analyze disease data and knowledge about senior citizens, who are in danger of contracting corona virus. The algorithms and tools can track these people right down to the postcode level, keeping in mind factors like obesity or diabetes. These analysis reports will suggest healthcare centres and hospitals where additional medical facilities like beds are going to be required.

Data scientists are performing on mobile applications which will be used for contact tracing. By utilizing the situation data on their smart phones, people are often alerted if they need been exposed to the virus. A team at Southern Illinois University (SIU) has developed a knowledge visualization tool that uses GPS data to alert users about locations of COVID-19 cases.

This data can be used and explore it by applying big data analytics techniques in different areas like diagnosis, estimate or predict risk score, healthcare decision-making, and pharmaceutical industry. [29]. Figure1shows examples of Major application areas of Big Data Analytics for Covid-19.

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



In the following subsections, we present several samples of COVID-19 data utilization from the literature with a primary specialise in reviewing studies that have provided solutions to regulate the COVID-19 pandemic and fall within one among the three areas, namely (1) diagnosis (Section2.1), (2) estimate or predict risk score (Section2.2), and (3) healthcare decision-making (Section 2.3).

2.1 DIAGNOSIS

Big data part in COVID-19 helped is becoming more noticeable as organizations like WHO, CDC and Microsoft are creating dashboards supported it. These virus tracking dashboards pull data from various countries and show number of confirmed cases, deaths and locations where it occurred. The dashboards are often wont to prepare datasets for giant data models. These giant data models can predict possible hotspots and warn the healthcare system beforehand only.

Another important big data process used against COVID-19 is outbreak analytics. This deals with the gathering and analysis of outbreak response data. Data including deaths, confirmed cases, tracing people contacted by infected patients, population densities, and far more are wont to develop data models for the disease. These data models can predict peak infection rates and their impact on patients.

Suspected cases of COVID-19 are diagnosed using the Reverse Transcription-Polymerase Chain Reaction (RT-PCR) test. This test takes around 24 hrs to many days, counting on the multiple conditions. Many countries experienced increased demand for diagnosing suspected cases of COVID-19, which exceeded the available local testing capability. There-fore, several researchers have proposed alternative solutions for the COVID-19 RT-PCR diagnosis test, including the subsequent.

The authors in Reference [30] have proposed a model to differentiate between COVID- 19 and 4 other viral chest diseases. The model utilizes several body sensors to gather information and monitor the patient's health condition, including temperature, vital sign, pulse, respiratory monitoring, glucose detection, et al. The gathered data is stored on a cloud database containing data related to AI-enabled expert systems that aid diagnosis.

Nose symptoms of patients suspected or infected of having COVID-19 to determine the correct procedure to deal with them. However, it's not clear how the patient's health information is going to be submitted to the hospital staff. Moreover, the authors in Reference [26] had presented various papers about mathematical models to extend the efficiency in detecting and predicting COVID-19. Their survey suggested using AI to detect COVID-19 cases, big data to trace cases, and nature-inspired computing (NIC) to pick suitable features to extend the accuracy of detection.

In Reference [31], the authors provided a versatile and low-cost design of a medical device which will be wont to detect and track symptoms of COVID-19. It utilizes headphones and a mobile to detect breathing problems. The signals are collected and saved in an audio file format through the mobile app, after which the signals are analyzed using the MATLAB program to spot the respiratory symptoms associated with COVID-19.

Researchers [32] also given a program to remotely observe discharged COVID-19 patients. For all Covid-19 patients registered to the app is issued a thermometer and pulse oximeter to self-examination daily symptoms, temperature and O2 saturation. The abnormal

vital signs and symptoms are flagged to be assessed by a gaggle of nurses. Depending on the evaluation outcome, the patient could be readmitted to the

Emergency Department (ED). The program helps to reduce utilization of Emergency Department and provides scalable remote monitoring capabilities when a patient gets discharged from the hospital.

The authors in Reference [33] found that smart watches might be utilized in COVID-19 pre-symptomatic detection. They examined the physiological and activity data collected from smart watches of the infected patients of COVID-19. They Observed, sixty-three percent of COVID-19 patient's cases could be found before symptoms appear by using a two-level warning system for severe elevations in resting pulse relative to individual baseline. Finally, they concluded that activity tracking and health monitoring using wearable devices can help in early detection of respiratory infections.

Since the COVID-19 symptoms haven't been fully identified and thanks to the changing nature of COVID-19, some studies have focused on identifying the medical characteristics and symptoms related to positive COVID-19 cases. The study in Reference [34] focused on identifying the symptoms related to the positive results of the COVID-19 examination, and it had been conducted on a gaggle of healthcare workers (HCWs). Initial screening was performed by phone, and a COVID-19 PCR test was also performed for every HCW to spot symptoms related to each case. The study found that the foremost common symptoms of positive COVID-19 cases were fever, myalgia, and anosmia/ageusia, while the negative cases mostly haven't any symptoms, or the symptoms are limited to nasal congestion and sore throat.

The study in Reference [35] aimed to determine the clinical characteristics and out- comes of 5700 hospitalized patients with COVID-19 in the NY area. However, the study included non-critically ill patients and therefore the follow-up time was limited.

Another study [36] proposed an internet site and Android app to separate a COVID-19 cough sound from other respiratory sounds with the help of crowd sourcing data from about 7000 unique users (more than 200 of whom reported a Covid-19 positive test). Their proposed method employed Logistic Regression (LR), Gradient Boosting Trees, and Support Vector Machines (SVMs) classifiers to differentiate the cough sound data supported gender, age, and symptoms. Also, their classifiers distinguish the user supported other features, like whether or not they are asthmatic patients, smokers, or healthy. Their app asks the user to cough from three to 5 times then repeat the method every two days to update the user's health status. Their method proved that a COVID-19 cough are often distinguished from other lung diseases coughs from the sound of the cough combined with breathing sound to screen the disorder. It attained 82% Area under the Curve (AUC) in identifying covid-19 cases that tested positive. They recommended more studies within the field to specify more characteristics of a COVID-19 cough sound to form it more distinguishable from other respiratory sounds.

The authors in Reference [37] announced the significance of using complementary technologies such as, on-body sensors for diagnosing and monitoring COVID-19 disease infections. They proved that clinical devices are more reliable and provide more functions than smart watches since these devices are kept in different areas of the physical body to find various body signals. A thin, soft sensor with precision temperature sensor and a high-bandwidth accelerometer placed on the neck is extremely crucial to record respiratory activities from cough frequency, intensity, and duration to rate of respiration and energy, to high-frequency respiratory features related to wheezing and sneezing. Also, they re-suggested machine learning and predictive algorithms to help to diagnose and monitor COVID-19 disease.

In Reference [38], researchers focused on the importance of detecting the characteristics of COVID-19 among patients of Saudi Arabia. The study involved 1519 cases where data associated with their genders, ages, vital signs, public data, and clinical examinations were gathered. Their test was conducted to supported the quantitative RT-PCR approach, which is the protocol established by the planet Health Organization. After the information was gathered, it had entered into electronic sheets with different data collectors, and entered data was analyzed with Statistical Package for social science programs.

The statistics manifested that the foremost 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 foremost infected patients' demographics include elder males, severe cardiac condition patients, and diabetic patients.

Thus, the authors in Reference [39] had proposed a computer virus method to assist the classification model to research the retinal image of diabetic retinopathy to research its effect among adults in causing blindness. It proved that the focused connection among layers of the convolution network helps the correctness of the classification result.

The retrospective, observational study in Reference [40] conducted a statistical analysis to point out the cardiovascular implications of COVID-19 on the patients. The study was done on 116 patients who tested positive for COVID-19. The info was clinically collected and tested to extract clinical symptoms and signs, chest computerized tomography, treatment measures, and medical records. The statistical analysis was done on the information to give similar results as those reported by Reference [38], where the common symptoms were fever and dry cough, and therefore the elder or middle-aged males, heart injury patients, hypertension patients, and diabetics were the first, most infected populations.

2.2 ESTIMATE OR PREDICT RISK SCORE

Big data can play a crucial role in analyzing screening data of patients and connecting it with unnamed health problems with hospitalized patients. The results of this analysis will definitely help in determining the key risk factors of covid-19. Hence, as more and more data is fed into big data programs and tools, the accuracy of risk predictions also increases.

Google's Deep Mind AI system takes support of big data to know the virus by examining its characteristics. This will help doctors to develop new drugs and treatment plans.

Estimating the danger score helps in determining the care level and priority for every patient with an insight to the required proactive measures. Within the following section, we present the studies that cover this area.

In Reference [41], the authors aimed to validate a hypothesis that COVID-19 infection could lead on to serious cardiovascular diseases or even worse. They used statistical analysis by involving a multi-factorial logistic regression model to perform research on COVID-19- related causes. The study was performed on 54 patients with different genders, ages, and vital signs, where 39 were diagnosed as severe cases of COVID-19 and 15 as critical cases of COVID-19.The information was collected clinically from the patients with attached sign measurement devices updated every four hours. Results showed that elder males, diabetic patients, and hypotension patients are more likely to develop a significant heart- related condition and wish more care. Their proposed study is restricted thanks to the tiny sample size, and that they suggested a better sample size to conduct a more appropriate study and verify the results.

The authors in Reference [42] have an interest in developing and validating the danger score to predict adverse events among patients suspected of getting COVID-19. They performed a retrospective cohort study of adult visits to the emergency department. The study concluded that the first outcome was death or no respiratory decomposition within 7 days. To derive the danger score, they used the smallest amount Absolute Shrinkage and Selection (LASSO) and Logistic Regression models. They concluded that the COVID-19 Acuity Score (COVAS) can help in decision-taking to discharge patients of COVID-19.They also stated the derivation and validation measures of cohorts and subgroups with pneumonia or COVID-19 diagnosis.

The authors in Reference [43] proposed an online of Things (IoT) based system to get unregistered COVID-19 patients, also as infectious places. This is able to help the responsible authorities to disinfect contaminated public places and quarantine the infected persons and their contacts albeit they didn't have any symptoms. The newly confirmed and recovered cases would be recorded within the system by the healthcare staff, while the geo-location data are going to be collected automatically by Global Positioning System (GPS) technology within the IoT devices. The authors discussed how their proposed system might be utilized to use three different prediction mathematical models, namely the

Susceptible-Infected-Recovered (SIR) model, θ-SEIHRD model, and Susceptible-Exposed- Infectious-Removed (SEIR) model.

Another study [44] demonstrated the likelihood of transmitting the COVID-19 virus through indirect contact, like touching surfaces contaminated with the droplets of an infected person. Therefore, it had been recommended that listening to non-public hygiene and disinfection of public places could possibly reduce the incidence.

Furthermore, researchers also [45] conducted a cross-sectional study to point out 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 needs more attention by the relevant authorities to deal with this impact. Also, the authors in Reference [46] had proposed a model that identified the danger of getting infected by tuberculosis supported several factors associated with tuberculin skin, age, and weak system. They stated that those factors can increase the infection from 10% to 20%.

The authors in Reference [47] presented a model that finds the course of the outbreak to help in planning an efficient method of prevention. SIDARTHE (susceptible, infected, diagnosed, ailing, recognized, threatened, healed, and extinct) Model Stages. It differentiates between infected people supported whether or not they are 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's an urgent necessity.

2.3 HEALTHCARE-DECISION MAKING

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

The authors in Reference [48] designed a prediction model called Conscious-based Susceptible-Exposed-Infective-Recovered (C-SEIR) model to make sure 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 therefore the quarantined diagnosed infection group (Q), alongside a blue/green curve with a solid line used for daily patients and dashed line used for cumulative patients. It showed that the results of the prediction may be a double drop-down or increase supported the town lockdown precautions in Wuhan. The authors also gave guidance for cover against COVID-19, like being educated about the virus, social distancing, and lockdown.

In Reference [49], the authors have developed a patient monitor program that permits daily electronic checking of symptoms, providing advice and reminders via text messages, and providing care by phone. Patients registered within the system complete a daily questionnaire to gauge 10 symptoms employing a scale from 0 to 4. Additionally to determining what proportion they feel the infection affects them, the amount of analgesic/antipyretic tablets they take, and therefore the temperature measured, questionnaire responses are wont to classify patients and specify the care needed. The study focused on three measures, namely the amount of patients monitored over time, the daily symptoms score, and daily ED referrals. similarly, the authors in Reference [50] invented a mobile app to find the spread of COVID-19 disease symptoms in UK by examining a group of information reported by patients registered within the application, including health risk factors, location, healthcare visits, age, symptoms and test results of COVID-19. This survey data helped them in determining patients' category and intensity,

availability of private protective equipment, and work-related stress and anxiety.

The study presented in Reference [51] was concerned with evaluating one among the COVID-19 applications in terms of user satisfaction and therefore the possibility of using the info collected to support decision-makers and healthcare providers. The app collects data daily from patients, including vital signs, symptoms, and an assessment of their feedback with the services provided by the app. the info collected is distributed on an interactive map consistent with the zip code for every user, which helps in knowing the

regional distribution of the spread of infection additionally to the share of healthcare consumption in each region.

Another study [52] presented an analytical model for estimating patient census and estimating ventilator requirement for a given hospital during the COVID-19 pandemic. Through this study, it had been noticed that the estimation of the bed and ventilator needs is influenced by the length of hospital stay, and therefore the number of days of inpatient ventilator use. Also, there was no relationship between the age of hospitalized patients and therefore the likelihood of needing a ventilator, or between the inpatient gender and therefore the length of stay. They recommended that every 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 therefore the number of deaths thanks to COVID-19 for subsequent four months from the date of the study for every state within the US [53].
The authors in Reference [54] tried to explain the clinical characteristics and identified factors that predict medical care unit (ICU) admission for COVID-19 patients. They found that the necessity for a COVID-19 patient to enter the ICU is often predicted by checking a group of medical parameters which will be easily obtained: age, fever, and tachypnea with/without respiratory crackles. They used the EHRead [55] model that was developed by Savana to retrieve data from the medical records. Also, deep learning convolution neural network classification methods are wont to classify the extracted data.
The authors in Reference [56] provided a data-driven framework to pre-assess the risks of the COVID-19 pandemic and to spot high-risk areas in Italy. The framework assesses the danger index employing a function consisting of three criteria, namely disease risk, area exposure, and therefore the vulnerability of its population. The twenty Italian regions are classified supported available historical data, which include population density, age, human mobility, pollution, and winter temperature. The study showed a correlation between the danger index and therefore the number of deaths, infected, and patients in ICU. They also provided a policy model to help authorities in making several decisions.
Moreover, regional healthcare models are developed to estimate the pandemic, just like the simulation approach developed at the University of Pennsylvania called Monte- Carlo [57]. Such models are often wont 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 a private hospital requires unknown parameters just like the proportion of the region's patients expected to go to that hospital, and therefore the percentage of the localized population isolated more to evade infection.

3. CONCLUSION

The volume of knowledge increases dramatically over time, especially data generated on the worldwide pandemic caused by COVID-19. Such volume of knowledge requires utilizing big data analytics tools alongside AI techniques to form sense of the pandemic and control its spread during a timely manner. during this study, we presented a review of several data analysis applications for COVID-19, providing a taxonomy structure which classified the potential applications of COVID-19 into four categories, namely diagnosis predict risk score, healthcare decision-making, and pharmaceutical. The paper introduced several data analysis tools and explained the most features of every tool. We also provided important insights on variety of challenges which may hinder the utilization of knowledge analytics tools for COVID-19. These challenges include healthcare data security and patient privacy issues, the problem of sharing data with researchers, absence of knowledge validation for a few studies which will cause biased results, and the patients support in sharing a part of their medical related information. Finally, we highlighted and discussed variety of future directions that ought to be considered in further research and applications to help stakeholders, like governments, MoHs, hospitals, patients, and responsible authorities, to make decisions and predict the future.

In the future, big data will definitely play a crucial role in analyzing global data about viruses detected, disease modelling, tracking and visualization of this disease data. As more and more data compile into huge datasets, data scientists will have a far better shot at preventing such outbreaks. Read more to find out how data science helps us to stop future pandemics.

4. FUTURE SCOPE

Big data role in COVID-19 – Securing our future.

After this global pandemic is resolved, big data can help the governments to stop and battle against future outbreaks. The information from this outbreak is often wont to test scenarios and analyze their outcomes to form vital decisions within the future.

REFERENCES

1. Perappadan, Bindu Shajan (30 January 2020). "India's first coronavirus infection confirmed in Kerala". The Hindu. ISSN 0971-751X. Retrieved 24 February 2021.
2. "India most infected by Covid-19 among Asian countries, leaves Turkey behind". Hindustan Times. 29 May 2020. Retrieved 30 May 2020.
3. "#IndiaFightsCorona COVID-19". MyGov.in. Govt of India. Retrieved 2 January 2021.
4.India's COVID crisis 'beyond heartbreaking': WHO". www.aljazeera.com. Retrieved 26 April 2021.
5."Infections over 1 lakh, five cities with half the cases: India's coronavirus story so far". The Week. Retrieved *20 May 2020.*
6. Shivani Kumar (10 June 2020). "Covid-19: Number of recoveries exceed active cases for first time". Hindustan Times. New Delhi. Retrieved 11 June 2020.
7. "With very high COVID-19 testing, India's positivity rate fallen below 8%: MoHFW". *The Economic Times. 18 October 2020. Retrieved 19 October 2020.*
8. "With very high COVID-19 testing, India's positivity rate fallen below 8%: MoHFW". The Economic Times. 18 October 2020. Retrieved 19 October 2020.
9.Michael Safi (21 April 2021). "India's shocking surge in Covid cases follows baffling decline". The Guardian. Retrieved 29 April 2021.
10. "Coronavirus | India becomes first country in the world to report over 4 lakh new cases on April 30, 2021". The Hindu. Special Correspondent. 30 April 2021. ISSN 0971-751X. Retrieved 2 May 2021.
11."India coronavirus: New record deaths as virus engulfs India". BBC News.
12. Ajah, I.A.; Nweke, H.F.; Ajah Big Data and Business Analytics: Trends, Platforms, Success Factors and Applications. *Big Data*
Cogn. Comput. **2019**, *3*, 32, doi:10.3390/bdcc3020032.
13. White, T. *Hadoop: The Definitive Guide*; 3rd ed.; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2012; ISBN 9781449311520.
14. 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, doi:10.1109/jbhi.2015.2450362.

15. 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 Mining: ASA Data Sci. J.* **2014**, *7*, 385–403, doi:10.1002/sam.11236.
16. 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, doi:10.1016/j.im.2017.04.001.
17. Wong, H.T.; Yin, Q.; Guo, Y.Q.; A. Murray, K.; Zhou, D.H.; Slade, D. Big Data as a New Approach in Emergency Medicine Research. *J. Acute Dis.* **2015**, *4*, 178–179, doi:10.1016/j.joad.2015.04.003.
18. Viceconti, M.; Hunter, P.; Hose, R. Big Data, Big Knowledge: Big Data for Personalized Healthcare. *IEEE J. Biomed. Health Inform.***2015**, *19*, 1209–1215, doi:10.1109/jbhi.2015.2406883.
19. 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, doi:10.1016/j.jmig.2015.07.003.
20. Nadkarni, G.N.; Coca, S.G.; Wyatt, C.M. Big Data in Nephrology: Promises and Pitfalls. *Kidney Int.* **2016**, *90*, 240–241, doi:10.1016/j.kint.2016.06.003.
21. Davis, R. Integrating Digital Technologies and Data-driven Telemedicine into Smart Healthcare during the COVID-19 Pandemic. *Am. J. Med Res.* **2020**, *7*, 22, doi:10.22381/ajmr7220203.
22. Valenzuela, S. Unpacking the Use of Social Media for Protest Behavior. *Am. Behav. Sci.* **2013**, *57*, 920–942, doi:10.1177/0002764213479375.
23. Sheng, J.; Amankwah-Amoah, J.; Wang, X.; Khan, Z. Managerial Responses to Online Reviews: A Text Analytics Approach. *BrJ. Manag.* **2019**, *30*, 315–327, doi:10.1111/1467-8551.12329.
24. Rehman, A.; Naz, S.; Razzak, M.I. Writer Identification using Machine Learning Approaches: A Comprehensive Review. *Multimedia Tools Appl.* **2019**, *78*, 10889–10931, doi:10.1007/s11042-018-6577-1.
25. 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, doi:10.1016/j.techfore.2015.12.019.
26. 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, doi:10.3390/ijerph17155330.
27. 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.; etal. Heart Failure: Preventing Disease and Death Worldwide. *ESC Hear. Fail.* **2014**, *1*, 4–25, doi:10.1002/ehf2.12005.
28. 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, doi:10.1016/j.dsx.2020.04.012.
29. PEX Process Excellence Network 6 Ways Pharmaceutical Companies are Using Big Data to Drive Innovation & Value. Available
online: https://www.processexcellencenetwork.com/tools-technologies/whitepapers/6-ways-pharmaceutical-companies-areusing-big-dat (accessed on 28 December 2020).
30. 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, doi:10.1177/1460458220952918.
31. 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); Institute of Electrical and Electronics Engineers (IEEE), Budva, Montenegro, July 2020; pp. 1–4.
32. 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, doi:10.1055/s-0040-1721039.
33. 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, doi:10.1038/s41551-020-00640-6.
34. 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, doi:10.1371/journal.pone.0235460.
35. 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, doi:10.1001/jama.2020.6775.
36. 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, June 2020; pp. 3474–3484.
37. Jeong, H.; Rogers, J.A.; Xu, S. Continuous on-Body Sensing for the COVID-19 Pandemic: Gaps and Opportunities. *Sci. Adv.* **2020**,
6, eabd4794, doi:10.1126/sciadv.abd4794.
38. Alsofayan, Y.M.; Althunayyan, S.M.; A. Khan, 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, doi:10.1016/j.jiph.2020.05.026.

39. Riaz, H.; Park, J.; Choi, H.; Kim, H.; Kim, J. Deep and Densely Connected Networks for Classification of Diabetic Retinopathy.
Diagn. **2020**, *10*, 24, doi:10.3390/diagnostics10010024.
40. 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, doi:10.1186/s12879-020-05452-2.
41. 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, doi:10.1002/clc.23384.
42. 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**, doi:10.1016/j.ajem.2020.10.068.
43. Benreguia, B.; Moumen, H.; Merzoug, M.A. Tracking COVID-19 by Tracking Infectious Trajectories. *IEEE Access* **2020**, *8*, 145242–
145255, doi:10.1109/access.2020.3015002.
44. 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, doi:10.1186/s12889-020-09296-y.
45. 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, doi:10.1016/j.jad.2020.07.135.
46. Horsburgh, C.R. Priorities for the Treatment of Latent Tuberculosis Infection in the United States. *New Engl. J. Med.* **2004**, *350*,
2060–2067, doi:10.1056/nejmsa031667.
47. 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, doi:10.1038/s41591-020-0883-7.
48. 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, doi:10.18562/ijee.055.
49. Kricke, G.; E.; Roemer, P.; Barnard, C.; Peipert, J.D.; Henschen, B.L.; A.; Bierman, J.; Blahnik, D.; Grant, M.; A. Linder, J. Rapid Implementation of an Outpatient Covid-19 Monitoring Program. *NEJM Catal. Innov. Care Deliv* **2020**, *1*, doi:10.1056/CAT.20.0214
50. 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,
doi:10.1126/science.abc0473.
51. 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, doi:10.2196/19822.
52. 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, doi:10.7759/cureus.8501.
53. 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**,
doi:https://doi.org/10.1101/2020.03.27.20043752.
54. 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, doi:10.2196/21801.
55. Medrano, I.H.; Guijarro, J.T.; Belda, C.; UreñaA.; 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, doi:10.9781/ijimai.2017.03.001.
56. Pluchino, A.; Biondo, A.E.; Giuffrida, N.; Inturri, G.; Latora, V.; Moli, R. Le; 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.
57. 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, doi:10.7326/m20-1260.