IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Data Analytics and AI help defeat COVID-19

Miss. Sneha Vilas Kotawadekar

Principal (I/C)
Computer Science,
Maharshi Karve Stree Shikshan Samstha's
College of Computer Applications for Women, Ratnagiri – 415 629 M.S. (India)
(Affiliated to SNDT University, Mumbai)

Abstract: Today corona virus outbreak has hit millions of lives across the world. And the rising threat of this virus is continuously increasing every day. Countries affected with corona virus are now taking major steps to address it using AI and big data technologies. The rapid, global spread of COVID-19 has brought advanced big data analytics tools to the front and center, where entities from all sectors of the healthcare industry are seeking to monitor and reduce the blow of this virus. This paper is an approach to determine the use of data analytics and AI to collect, analyze and predict different models, equations etc of the outbreak that help to fight against novel corona virus.

Index Terms **– COVID-19, RIP, BlueDot, NLP, SIR**

I. INTRODUCTION

Today almost whole world is in the state of crises because of the rising threat of COVID-19.Between recognizing signs and symptoms, tracking the virus, and monitoring the availability of hospital resources, researchers are dealing with enormous amounts of information which is too much for humans to seize and analyze on their own. It's a situation that is seemingly custom-made for advanced analytics technologies.

There are several big data components to this pandemic where artificial intelligence can play a big role. One component is biomedical research. A lot of work is going on to try to develop a vaccine to find out whether there are any current drugs work against COVID-19. All of those projects require molecular modeling, and many of them are using AI and machine learning to map things we know about the virus to things in pharmacological databases and genomic databases. Several big-name organizations have launched projects like these – Amazon Web Services, Google Cloud, and others have recently offered researchers free access to open datasets and analytics tools to help them develop COVID-19 solutions faster.AI can eliminate many false tracks and allow us to identify potential targets. So instead of trying 100 or 1000 different things, we can we narrow it down to a much smaller size much faster. That's going to accelerate the eventual finding of the vaccine. Researchers are also leveraging AI to evaluate the effects of COVID-19 interventions on individuals across the country [2].

According to the World Health Organization (WHO), AI and big data played a significant role in China's response to COVID-19, new name of corona virus [1].China has millions of security cameras that are used to track citizens' movements in a view of spotting crimes. This helped authorities discover people who weren't compliant with quarantine orders. If a person was supposed to be in quarantine, but cameras tracked them outside their homes, authorities would be called. Mobile phone data was also used to track movements. The Chinese government also used a "Close Contact Detector" app that alerted users if they were in contact with someone who had the virus. Travel verification reports produced by telecom providers could list all the cities visited by a user in the last 14 days to determine if quarantine was recommended based on their locations. By integrating the data collected by China's surveillance system, the country was able to find ways to fight the spread of the corona virus.

At Rensselaer Polytechnic Institute (RPI) researchers are using big data and analytics to better comprehend corona virus from a number of different angles. The institute recently announced that it would offer government entities, research organizations, and industry access to innovative AI tools, as well as experts in data and public health to help combat COVID-19.The institute is working with several organizations on modeling and dealing with the virus directly using a supercomputer, and created some websites where a track is kept of all the open data and documents that one can find to help researchers find what they're looking for. The institute is also working to understand social media responses to the pandemic. One project in particular has focused on tracking data from Chinese social media as corona virus spread there in mid-January, and then comparing it to American data [2].

A team of experts at Oxford University said their research demonstrated the viability of a smart phone app to provide instant contact tracing—the ability to identify people who have had contact with confirmed Covid-19 cases—that would significantly help to contain the spread of corona virus. According to Christopher Fraser, a professor of pathogen dynamics at Oxford University's Big Data Institute, who led the team, the app could be used as a key data source for outbreak analytics [3] .

I.1 Method:

The entire discussion in the paper is based on the secondary data. The sources used are duly cited in the text of the paper and detail references of the same are given at the end in the bibliography.

I.2 Theoretical framework:

In 2014, the Centers for Disease Control (CDC) predicted up to 1.4 million cases infected from the Ebola outbreak. Then, it was said that what big data could do to help identify the earlier signs of future outbreaks. That time, Harvard's HealthMap service made headlines for monitoring early mentions of the Ebola outbreak nine days before the WHO officially announced the epidemic, and issuing its first alert. Reports suggest that the HealthMap's early warning came from using massive computing power to sift out early indicators from millions of social media posts and other media platforms. But an article published on FP, "Why Big Data Missed the Early Warning Signs of Ebola", mentioned that by the time HealthMap monitored its very first report, the Guinean government had already announced the outbreak and notified the WHO .Now, in the case of coro na virus outbreak, On December 30, 2019, a Toronto-based startup BlueDot that uses a platform built around AI, machine learning and big data to track and envisage the outbreak and spread of infectious diseases, alerted its private sector and government clients about a cluster of unusual pneumonia cases happening around a market in Wuhan, China. In this case, some reports earlier claimed 27 pneumonia cases associated with a market that had seafood and live animals in Wuhan. And the cities BlueDot identified was correct that were highly connected to Wuhan using things like global airline ticketing data to help anticipate where the infected might be traveling. The international destinations that BlueDot anticipated would have the highest volume of travelers from Wuhan were: Bangkok, Hong Kong, Tokyo, Taipei, Phuket, Seoul, and Singapore, according to an article. In the end, 11 of the cities at the top of their list were the first places to see COVID-19 cases.

But looking at the infection rate graph, the 11 cities Blue Dot listed are near to flat on the bottom, and countries like South Korea, France, Germany and the US are climbing upwards [1] .

Epidemiologists are teaming up with data scientists to stem the spread of the novel corona virus by tapping big data, machine learning and other digital tools. The goal is to get real-time forecasts and other critical information to front-line health-care workers and public policy makers as the outbreak unfolds. Known as outbreak analytics, the approach attempts to go beyond confirmed cases and fatalities to gauge the total number of people who are likely to be infected by a virus—whether or not they show any symptoms. Besides advancements in medicine, advancements in information technology and digital data are how we defeat this pandemic and prevent another Spanish flu-like outcome. Outbreak analytics seeks to gather all available data on an epidemic, including confirmed cases, fatalities, test results, tracing contacts of infected people, maps of population densities and demographics, traveler flows and migration, availability of health-care services, drug stockpiles and other factors. The raw data is then processed into compatible formats—increasingly by machine-learning software trained to recognize patterns and clean up diverse data sources—and fed into algorithmic models. The models are designed to predict the number of new cases that are likely to arise in an exposed population, or peak infection rates under a given set of conditions, among other outcomes [3].

I.3 Statistical tools:

I.3.1 Descriptive Statistics:

With SARS and other outbreaks, we never really had to figure out how different social distancing techniques are impacting the spread in different places. AI is very good at that kind of multi-factor learning and a lot of people are trying to apply those techniques now. At UTHealth, a team developed an AI tool that showed the need for stricter, immediate interventions in the Greater Houston area. And at Stanford University, researchers have launched a data-driven model that predicts possible outcomes of various intervention strategies .Using big data and analytics tools of their own, Hendler and team of this university are aiming to do something similar .There is a lot of time series data from China, there is information about airline transportation, and population models for each country. Now they are looking at doing this in their own region, and seeing if they can track and predict the spread based on the kind of social measures taken within different regions.AI can also help organizations draw on research from the past, applying this knowledge to present and future situations . Another area where AI can make an impact is in mining scientific literature. AI and other analytics technologies appear to be the best possible tools for assessing and mitigating a global pandemic, researchers can't always access what they need to build these models .The ideal data is hospital data that would tell us who is experiencing certain impacts from the virus .Information about how people are moving, the effect of travel restrictions or stay at home orders, how many people have what – that's data we can get. The unprecedented impact of corona virus around the world has sparked the need for unprecedented partnerships, and these collaborations will contribute significantly to finding viable solutions. In healthcare, academia and industry are mostly set up for people to stay in their own lanes. But people are rapidly beginning to realize that attacking this problem is going to require a collaborative effort. To make any real progress in this situation, you need to bring together people who understand the computation and AI, people who understand the biological and biomedical implications, and people who understand population models [2] .

I.3.2. Datasets and Applications of Machine Learning to the Corona virus:

1) Most applications fall under one of four areas:

(A) Predict the structure of proteins and their interactions with chemical compounds to facilitate new antiviral drugs/vaccines or recommend current drugs .Methods here rely on applying deep learning to molecules such as proteins. This is a somewhat niche area that generally has a high learning curve to understand. However, breakthroughs here could potentially pave the way to vaccines or an effective antiviral. A majority of the techniques listed below use convolution neural networks in some way to model molecules or molecular interactions.

1. Deep Learning Based Drug Screening for Novel Corona virus 2019-nCov (Zhang, et al.). The authors use a modified DenseNet (with the convolution replaced by a fully connected layer) to predict protein-ligand interactions. They can then use the model with the RNA sequences of corona virus together with chemical compounds to predict which drugs work the best. The authors conclude that more research is necessary, but suggest that Adenosine, Vidabrine, and other compounds might potentially help.

2. Predicting commercially available antiviral drugs that may act on the novel corona virus (2019-nCoV), Wuhan, China through a drug-target interaction deep learning model .This is similar to the paper discussed above, however the authors look at commercially available drugs and take a completely different modeling approach. Here the authors use a network named the "Molecule transformer-drug target" or MT-DTI. Fascinatingly, as those familiar with BERT might have guessed this at its core is the same architecture. However, in this case the network was trained on the SMILES dataset, a large dataset that represents molecules as text to encode and decode each molecule. This can actually form effective representations of molecules in much the same way it does for textual data. The authors then fine-tuned this pre-trained model to predict "binding affinity values between commercially available antiviral drugs and target proteins." They found that "the 2019-nCoV 3C-like proteinase was predicted to bind with atazanavir." Specifically, atazanavir is an anti-viral medication used to treat HIV/AIDS. This provides a good use case both of how new deep learning architectures in different domains can be adapted (though obviously this should not be taken as medical advice).

3. Deepmind used available data from GISAid and their AlphaFold library to predict the protein structures of Covid-19 virus. AlphaFold is a deep learning library for computational chemistry. With these protein structures (if correct), researchers will gain insights into the molecular structure of the virus which could potentially pave the way for finding vaccines or antiviral quicker.

(B) Forecast infection rates and spread/patient prognosis to enable hospitals/health officials to better plan resourcing and response. There actually haven't been too many models (at least publicly documented ones) that have explicitly attempted to model the corona virus spread. However significant amounts of prior research have studied forecasting the seasonal flu and other outbreaks. Interestingly, a large number of the methods currently being used to forecast disease spread and patient mortality are based on shallow methods .There is a lot of potential to use deep learning models with attention and transfer learning (on related flu outbreak data) to achieve better results in this area.

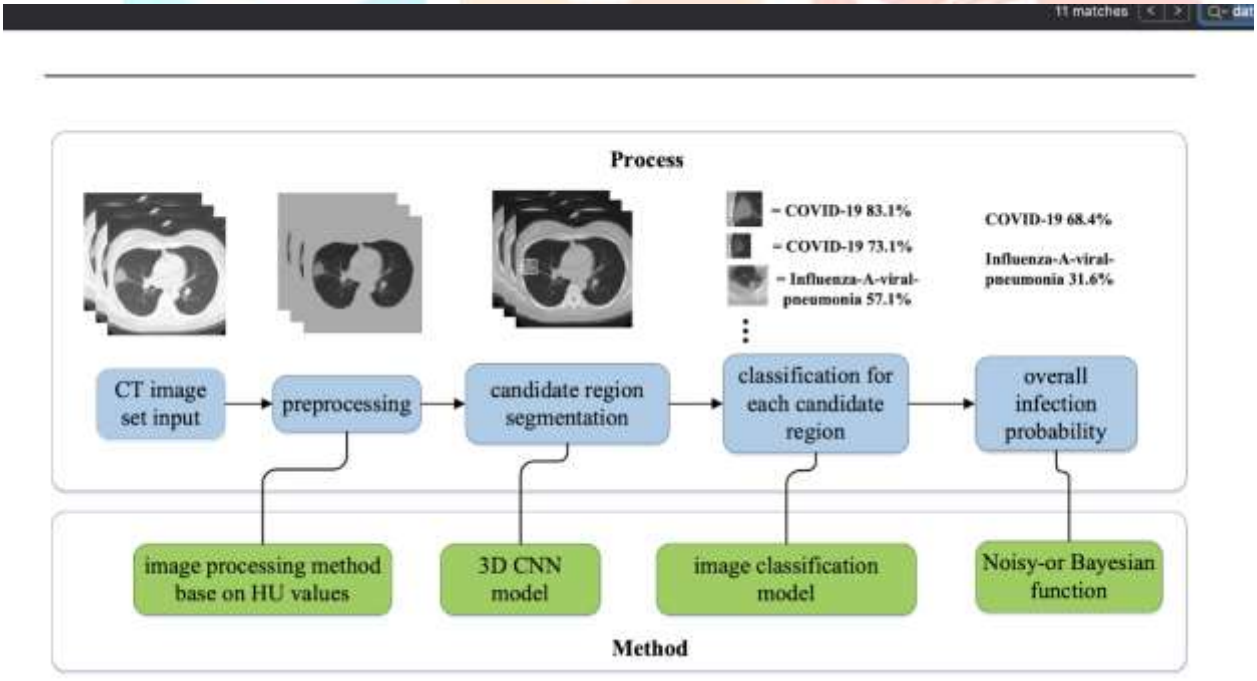
1. Prediction of criticality in patients with severe Covid-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in Wuhan: In this article, the authors describe using a XG-Boost model to predict if a patient infected with Covid-19 would survive the infection based on age and other risk factors. This is useful for forming recommendations about who should isolate themselves from the disease the most.

(C) Help diagnose if a medical image like a X-Ray or CT shows corona virus. Diagnosing a case of a corona virus related pneumonia from CT scan can, potentially, both shorten the time of diagnosis and enable better treatment. With record numbers of patients flooding ICUs, radiologists can quickly become overwhelmed. Deep learning on imaging can help lessen the burden. Moreover, learning how disease manifests itself in CT-scans can help provide more insight into the disease itself.

1. Deep learning-based model for detecting 2019 novel corona virus pneumonia on high-resolution computed tomography: a prospective study (Chen et al) Here the authors used a UNet++ to extract relevant features from the CT scans and classify them. The authors trained the model on 40k scans from 106 admitted patients. The authors found that the model performed well in terms of precision and recall. Specifically: "The model achieved a per-patient sensitivity of 100%, specificity of 93.55%, accuracy of 95.24%, PPV of 84.62%, and NPV of 100%; a per-image sensitivity of 94.34%, specificity of 99.16%, accuracy of 98.85%, PPV of 88.37%, and NPV of 99.61% in retrospective dataset. For 27 prospective patients, the model achieved a comparable performance to that of an expert radiologist. With the assistance of the model, the reading time of radiologists was greatly decreased by 65%." So, in summary, the model was highly effective at optimizing the workflow of radiologists. This is a good example of a positive impact of AI in real-world cases. Thus radiologists can examine more CTs in the same amount of time. It is important to note, however, the exceptional scores seemed to be on a test set of only 27 patients for prospective testing and 44 patients for retrospective testing. Moreover, all of these patients were from the same hospital. While this shows the model learned something it is unclear how robust it would be on a larger and more diverse test population.

2. A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19)

Figure 1. From paper a deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19)



(D) Mine social media data to better estimate spread/symptoms and general public perception:

This area of work focuses on mining social media data for relevant information about the disease. Although social media is very noisy in some ways it could contain more information about symptoms/spread in the general public. Once again at this point, no one (at least openly) has conducted research on mining social media data explicitly for Corona virus tracking [4].

However predictive analytics and big data have become a powerful tool for combating deadly disease outbreaks such as corona virus and SARS.

2) BlueDot:

BlueDot is one AI start-up that has developed intelligent systems that sift through data about people to determine the chances of disease occurrence. The AI platform from BlueDot is among the latest technological advances using data analytics to map and prevent diseases. Something quite interesting is that BlueDot predicted the SARS pandemic and turned out to be true. The SARS outbreak came with devastating effects and claimed the lives of almost 1,000 people. The outbreak alert about the Corona virus on December 2019 is another evidence attesting to the powerful nature of AI technology.

3) Natural language processing (NLP):

NLP is one tool used by BlueDot to track diseases with the company being successful in detecting diseases around the globe. For instance, BlueDot analyzes human languages around the world and use the information to assist them forecast disease outbreaks. Machine learning is another technology used by BlueDot with the algorithms offering updated information about possible disease occurrences. Nevertheless, the #AI technology at BlueDot saves time and resources by empowering health professionals with information on prevention measures. Often times, disease prevention poses threats compared to predicting and thanks to the machine learning tools, health specialists focus on patient safety. For example, NLP is aiding in distinguishing between people complaining about corona virus symptoms and those discussing about the disease but not affected. According to the HMS, the patterns in online media can facilitate discovery of a location outbreak and promote increased awareness on potential solutions.

The University of Southampton also is undertaking research efforts in the current corona virus outbreak with the institution using AI technology to model data from search engines to map the outbreak. According to the researchers, AI technology is assisting them to understand the movement patterns of the corona virus from Wuhan to other parts of China and the rest of the world. These machine learning and AI technologies have assisted researchers to predict the virus, its structure and its spreading methods. Consequently, this will help health professionals understand the solutions needed to combat further spread of the virus.

4) Geographic Information Systems Technology:

The GIS technology has become an important tool for stopping the spread of the coronavirus with John Hopkins University leading the way in this area. For example, the institution has a dashboard that shows all cases of corona virus around the world

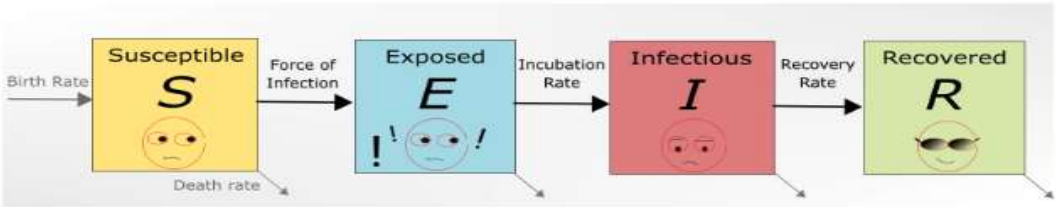
Data mining is critical for GIS technology to work because of using information to detect areas where people talk about the disease. Social media sites are good information sources for GIS as the technology maps the area of interest where people are talking about the corona virus. Accordingly, prevention measures can be implemented since these heatmaps can better track both the location and the spread of a disease. Ten years ago, it was practically impossible to track diseases; today, with AI, machine learning and GIS, data mining and extracting insights is both easier and more powerful at location viruses. The bottom line: prevention response time is quicker today [5] .

5) SIR models:

Running models through data analysis systems has proven to be able to approximate how trends might progress. An example is the SIR model; it is an epidemiological model that computes the theoretical number of people infected with a contagious illness in a closed population over time. The model uses coupled equations analyzing the number of susceptible people S(t), number of people infected I(t), and number of people who have recovered R(t). One of the simplest SIR models is the Kermack-McKendrick model. The Kermack-McKendrick epidemic model is considered the foundation on which many other compartmental models were based. R0 is an average, so it can be influenced by factors such as super-spreaders events. Tracking each case and the transmission of the disease is extremely difficult, so estimating R0 is complex and challenging.

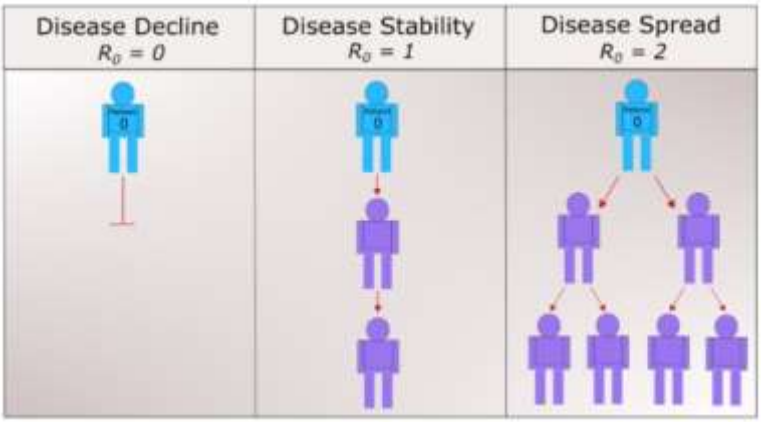
According to Ettore Mariotti's analysis it considers an island, our system, where people are not allowed in or out. Every individual can be in one of the following states at a given time: "Susceptible," "Infected" and "Recovered," hence the acronym SIR, because with a certain probability people who have never had the disease (S) can become ill and infected (I) for a certain period before they recover (R). In the case of CoVID-19, it is appropriate to extend the model with an additional state, "Exposed," to include people who have the virus but are not yet infectious (SEIR model).

Figure 2: *SEIR Model*



This model considers two factors: the dynamics of the virus, and the interaction of individuals. The latter is very complex and would require technology like the one described in the previous paragraph. With all this, it is possible to define the R0 parameter, which represents the number of people that an infected person can potentially infect. for example, that person A is sick and that our system has an R0=2. This would mean that A will infect two people. Those two people will, in turn, infect four people, who will infect another two people each (so 4 * 2 = 8) and so on. This highlights the fact that the spread of the disease is multiplicative rather than additive. R0 can capture three basic scenarios, as shown in Figure 3.

Figure 3. *R0 basic scenarios*



The closure of schools, gyms, etc. decreases the social interaction of people, thus lowering R0. The health system is limited, and it is very important to reduce this parameter below unity. If R0>1 the disease spreads, if R0<1, the disease disappears. It is reasonable to expect governments to impose stricter constraints on people's mobility in an attempt to reduce R0. R0 measures the potential transmission of a disease, not the rate at which the disease spreads [6] .

RESULTS AND DISCUSSION

There are countless approaches from universities, Institutes of all over the world trying to put forth their view, equations and models to help predict how to fight spread of corona virus. Topmost many highpowered AI companies like **Blue Dot**: Bluedot is one of the first companies that sensed the upcoming healthcare crisis in December, even before the WHO. The company on December 30, analyzed a cluster of "unusual pneumonia" cases happening around a market in Wuhan, China, and flagged it. The key to BlueDot is big data. It uses natural language processing and machine learning to cull data from hundreds of thousands of sources, including statements from official public health organizations, digital media, global airline ticketing data, livestock health reports and population demographics, **Deargen**: According to its prediction, from available FDA-approved antiviral drugs, the HIV medication atazanavir is highly expected to bind and block a prominent protein on the outside of SARS-CoV-2, the virus that causes COVID-19. The model also detected three other antivirals that can possibly bind the virus. , **Baidu:** Due to the recent outbreak, Baidu AI scientists have used this algorithm to predict the secondary structure prediction for the Covid-19 RNA sequence, reducing overall analysis time from 55 minutes to 27 seconds, meaning it is 120 times faster. , **Insilico Medicine** : Insilico says it will synthesize and test 100 of the most promising candidates while publishing the full library of new molecular structures it has generated for other researchers to possibly use., **SRIBiosciences and iktos:** Iktos, a company specialized in artificial intelligence (AI) for novel drug design and SRI International (SRI), a research center headquartered in Menlo Park, California, recently announced that the companies have entered into a collaboration agreement designed to accelerate discovery and development of novel antiviral therapies. Under the collaboration, Iktos' generative modeling technology will be combined with SRI's SynFini™, a fully automated end-to-end synthetic chemistry system, to design novel, optimized compounds and accelerate the identification of drug candidates to treat multiple viruses, including influenza and the corona virus (COVID-19)[7] .

Almost all are trying their best to come up with more exact model and solution. It is very obvious to believe that without AI and data analytics human being on today's date would not be able to help other human as today they can. All its efficient tools and brains are working hard day and night to reach next level of height everyday so that data processed here can help COVID worriers like Medical Practitioners to cope up with this challenge. There are many more facets of the pandemic where AI and data analytics is turning out to be a boon to visualizes, manipulate and model corona virus data which can eventually help to provide a proper vaccine in near future.

V. ACKNOWLEDGEMENT

The author is grateful to the individual researchers and the organizations those who have contributed scholarly on to the topic "Data Analytics and AI help defeat COVID-19" and made their contributions available to others. They have brought the subject in the lime light with various aspects of it. The author has given due credit to all those concern and also cited their literature. The author is thankful to all the Authorities of Maharshi Karve Stree Shikshan Sanstha, Pune and Ratnagiri. Thanks are also expressed to Mr. R. G. Sawant, Asst. Prof., Department of Commerce of R. P. Gogate College of Arts & Science and R. V. Jogalekar College of Commerce, Ratnagiri for his continuous support and valuable guidance from time to time in writing this paper.

REFERENCES

[1] Vivek Kumar ,(2020).*IS BIG DATA EFFECTIVE IN RESPONSE TO CORONAVIRUS OUTBREAK ?*
Retrieved from https://www.analyticsinsight.net/big-data-effective-response-coronavirus-outbreak/ . Accessed on:1st April 2020.
[2] Jessica Kent,(2020). *Understanding the COVID-19 Pandemic as a Big Data Analytics Issue.*
Retrieved from https://healthitanalytics.com/news/understanding-the-covid-19-pandemic-as-a-big-data-analytics-issue. Accessed on:4th April 2020.s
[3] Angus Loten,(2020). *Scientists Crunch Data to Predict How Many People Will Get Coronavirus.*
Retrieved from https://www.wsj.com/articles/ .Accessed on 4th April 2020.
[4] Isaac Godfried (2020). *Machine Learning methods to aid in Coronavirus Response.*
Retrieved from https://towardsdatascience.com/machine-learning-methods-to-aid-in-coronavirus-response.Accessed on 4th April 2020.
[5] David Yakobovitch(2020).*How to Fight the Coronavirus with AI and Data Science* .
Retrieved from https://towardsdatascience.com/how-to-fight-the-coronavirus-with-ai-and-data-science-b3b701f8a08a.
Accessed on 10th April 2020.
[6] M. Di Paolo Emilio (2020). *Covid-19: The Power of AI and Big Data*.
Retrieved from https://www.eetindia.co.in/covid-19-the-power-of-ai-and-big-data.Accessed on 10th April 2020.
[7] Smriti Srivastava, *Analytics Insight , (March 2020)*TOP 10 AI POWERED COMPANIES STANDING AGAINST CORONAVIRUS PANDEMIC. Retrieved from https://www.analyticsinsight.net/top-10-ai-powered-companies-standing-coronavirus-pandemic/ Accessed on 10th April 2020