Healthcare Analytics: An Advent to Mitigate the Risks and Impacts of a Pandemic

Shubham Kumar

Abstract Healthcare analytics is a broad term for a specific facet of analytics which sweeps in a wide sash of the healthcare industry, providing macro- and micro-level insights on patient data, risk scoring for chronic diseases, hospital management, optimizing costs, expediting diagnosis and so on. An efficacious way of implementing healthcare analytics on this sector is to combine predictive analysis, data visualization tools and business suites to get valuable information that can both deliver actionable insights to support decisions, as well as reduce variations to optimize utilization. In order to device a more potent approach to prepare for the next pandemic, the clinical data has to be improved which would provide more granulated information to augment treatment effectiveness and success rate. Confirmatory data analysis combined with predictive modeling can be used to furnish unexpected gaps in terms of getting the required clinical data, which is vital for the success of this approach. The healthcare analytics domain has got the potential to monitor the health of the masses to identify disease trends and can provided enhanced health strategies based on demographics, geography and socio-economics. Hence, healthcare analytics is indeed an advent to counter the imminent threats of outbreaks such as epidemics and pandemics provided it gets the necessary course of action to fulfill its grail.

Keywords Healthcare analytics · Big data · Data analytics · Business intelligence · Seaborn

1 Introduction

1.1 Healthcare Sector

Making itself as the prima facie of the human development the healthcare domain is certainly one of the largest and most prominent industries around the globe. This

S. Kumar (✉)
Deloitte Consulting India Private Limited, Hyderabad, India
e-mail: shubham.sbkr@gmail.com

175

industry goes in tandem with the survival lifestyle and living quality of the modern humans and hence reflects a direct reverberation of people living in different continents, countries or cities of the world. If one has to define it, healthcare is the diagnosis, cure and prevention of any mental or physical impairments such as an injury, disease or illness. It is performed, grafted and delivered by trained practitioners who could be present in different domains across medicine, psychology, physiotherapy, dentistry, chiropractic, nursing, pharmacy, allied heath and other care providers. This industry is responsible for end-to-end steering of the medical procedures which ensure the correct treatment of patients. This sector has been present in one or the other forms, since ancient times and today, it is the fastest growing sector in the world. The delivery of this sector depends on how the internal segregated group of professionals' function while being aligned together. The healthcare industry also interreacts with different domains to provide augmented and optimized services.

1.2 Analytics Domain

If one has to define analytics, it can be expounded as the application of a progression of phases, which are marked by algorithms or transformations, to generate cognizance from a processed set of datasets. It can be further segregated into in process of fetching, processing, analyzing and interpreting data to gain more knowledge or insight about it. In recent times, especially after the invariable expansion of stored data, analytics has gained a paramount status across multiple domains, sectors or industries. From gaining insights about the nature of crime and criminal psychology to boosting the sales figures of a supermarket, this technology has penetrated across myriad sectors and is considered predominantly utilitarian. It is gaining more prevalence due to the advancement of the quantum computing and storage technology. In the previous generation of information technology, the processing of data was slow and there were limited space to store a set of data. This was a major hurdle in the growth of analytics sector. But, with the augmented ability of machines and systems to store, read and process humongous amount of data in a very less period of time, that hurdle has been crossed. Though there could be a myriad solutions coming out of the analytics domain, it is mostly segregated into three categories which are—descriptive analytics, predictive analytics and prescriptive analytics. At times, the combination of one or more than one of these categories is used to provide an appropriate solution. The first category, descriptive analytics, tries to lay down the already present fact which has been hidden in the vast layers of a dataset. Here, the already captured data is just altered in dimension to get to the internal layer of information which is aggregated as a commodity. The function of predictive analytics starts where the descriptive analysis ended. It uses the results provided by the descriptive analytics as a base to try and predict the future or imminent events. It does not give an exact description of a future event but merely provides an idea of what could be imminent. The final category, prescriptive analytics uses both the descriptive and predictive

analytics to suggest the next course of action depending upon the outcomes of a stimulating various potential scenarios.

1.3 Application of Analytics in Healthcare Domain

Analytics has a ubiquitous impact on almost all the industries and healthcare is a sweeping example. To express in a nutshell, analytics has left an ineffaceable imprint on the healthcare industry. There are many live examples which bolster this affirmation, such as the omnipresence of fitness bands and the optimized use of the incoming data and information. This piece of technology has indeed revolutionized healthcare and changed the approach of dealing with diseases, physiological, outbreaks, epidemics and even a pandemic. All sectors across the society are benefitting from this gigantic volume of healthcare data which include the private and the public sector. Such a synergistic domain presents a colossal elbow room for professionals across different sectors to learn this component of advanced analytics, in order to transmogrify business challenges into business successes.

2 Background

Mass health casualty has always been a worrying topic for mankind and there are various accounts of the battle of survival between humans and diseases, which took the form of an epidemic or pandemic. Since Neanderthal's times, humans have fought and survived various ailments of varying potency. Throughout the extensive era of humans, they have been ravaged by the sporadic but ferocious presence of plagues and epidemics which in some scenarios, changed the course of history. The oldest account of a prehistoric pandemic is of circa 3000 B.C when an epidemic completed eradicated a prehistoric village in China. Currently known as "Hamin Mangha," this prehistoric site is among the most primitive and most preserved across the globe [1]. The current records based on a lot of anthropological study stipulate that the epidemic engulfed the population so quickly that there was no time for the proper burial of people who died. Similarly, the plague of Athens is also considered one of the gravest in the history. This epidemic started just after a huge war started between Athens and Sparta which lasted for five long years. According to a lot of studies, this plague is said to engulf as many as a hundred thousand people which was a sizable chunk of population, back in 430 B. C. In the modern era, The Black Death which traveled from Asia to Europe between 1346 and 1353 was among the worst pandemics mankind have face [1]. According to some studies, this even wiped out half of the population across Europe. Some of the other examples for epidemics which succeeded these events were the London Plague, yellow fever, etc. The last pandemic of world which was in presence, volume, effect and destruction comparable to COVID-19, was the Spanish Flu of 1918. It is estimated 500 million

people fell victim to this pandemic [1]. The people were from almost all parts of the globe ranging from the South Seas to the North Pole. It is estimated that one-fifth of the population who came in contact of this diseases ultimately died. Because of its potency, some indigenous communities were shoved to the verge of extinction. The unfurling and destructiveness of this pandemic was augmented by the confined and restricted conditions of soldiers during World War I. The poor nutrition which was a byproduct of war also contributed to its lethality. There is an interesting story about the name of this disease. Contrary to popular belief, the outbreak did not start in Spain. During that era, Spain had minimum censorship for press and reporting which resulted into the open publication of the accounts of this outbreak, and consecutive circulation in the Spanish newspapers and magazines. This let to people erroneously believing that it originated from Spain, and hence, the name Spanish Flu became prominent [1].

Talking about the current pandemic, COVID-19 (Coronavirus disease 2019) is a viral infectious disease which allegedly started in December 2019 in the Wuhan district of China. It is caused by SARS-CoV-2 and people suffering from it get flu like symptoms along with breathlessness which is usually an indication of lung infection [2, 3]. In 2020, the World Health Organization (WHO) declared COVID-19 as a pandemic. As of September 2020, over 30 million people had been infected globally with over 1 million deaths. There is no recognized account of the origin of this disease while a lot of studies indicate that it got transferred from bats in Wuhan Market of the Hubei province of China. This disease spreads from close proximity and surface transmission, though there are proven studies which indicate the spread though air transmission [2]. This disease sent the world into a state of panic and hysteria. Mass lockdowns were initiated in different cities across the globe along with the practice of social distancing and quarantine, to curtail the spread of this disease. As of now, the SARSr-CoV virion causing this disease has been observed to be destroyed by the application of a household soap which bursts it protective bubble, provided it is present outside a human body [2, 3].

3   Research on Pandemics and Their Impacts

Once the definition of a pandemic is delineated, then comes the part where its morbidity is accessed. Pandemics are essentially widespread, general cases of global outbreaks of infectious ailment which have multifaceted effects on one or more essential aspect of human survival. They are capable of causing sudden and imminent disruption across innumerable fields such as health, economics, and politics and so on. The impact usually lasts for years and the recovery takes an extra amount of physical and physiological resources. In the modern era, the likelihood of the adverse effects of a pandemic has increased by multiple folds. The prime reason for this change is the augmented and optimized travel options present in the modern era which acts as the peripheral carrier to the disease. Medical science has to change its approach while dealing with a natural hazard of this dimension. Also, specific policies

were brought into mitigate the effect of nascent outbreaks which had the potential to become a pandemic. Also, a decent amount of effort was put on to expand the sector which would work toward the sustainable development in creating the preparedness and heath capacity for an imminent outbreak. In the wake of the current century, the world had to deal with an outbreak which had the potency to become a pandemic, namely Severe Acute Respiratory Syndrome (SARS) [4]. In this case, a huge setback was received in the form of delayed reporting which directly contradicts the whole purpose of having a goal of being prepared. The sheer damage caused by this outbreak in a very short span of time forced the World Health Organization to bring major changes to the International Health Regulations (IHR) [4]. Here, the substantial necessity molded innovation and the changes helped in rapid yet outstretched testing of the affected individuals and potential targets. The specific standards delineated for detecting, responding and reporting helped the medical community in being prepared for the imminent danger. The world achieved a short-lived success later in the decade when the effects of 2009 influenza pandemic were successfully mollified with the help of rapid response and planning ahead of time. But, this triumph was short-lived when the next decade exposed the shortcomings of International Health Regulations (IHF). The significant aperture and slits were exposed when the world faced many such outbreaks of similar nature, one example of which can be the 2014 Ebola Epidemic. It was now evidently clear that there were significant challenges for proper screening and detection of the ailment, availability of isolation facility, basic care facility for the affected masses, tracing of potential contacts, coordination, mobilization and so on and so forth [4].

In order to be ready for the next pandemic, the risks, impacts and mitigation have to be delineated. A separate listicle has to be maintained for knowledge and technological gaps which would help the mankind to be prepared and equipped both physically and psychologically to combat an outbreak. Here, analytics plays a crucial role in gathering and asserting the required information on the above three verticals. There are a number of steps which has to be implemented to fully understand the potency of the risks involved in a potential pandemic outbreak. It is a well-known fact that pandemics have occurred often in the history, but in the current epoch, it is more likely to happen because of the surging exposure of viral disease particularly from animals. If one had to segregate the types of risks involved in an outbreak, then it would be broadly classified into two categories—spark and spread. The first category denotes the instantaneous risk at the time when an outbreak comes into existence. Whereas, the second category denotes the risk when an outbreak tries to unfurl and outspread, thus taking a form of an epidemic or in worst case, a pandemic. The population across the globe is not uniform in terms of access to healthcare and exposure to virion. A common myth was exalted till the year 2020 that Central and West Africa are at high instantaneous risk and they do not possess a steady amount of preparedness when compared to the rest of the world. This myth was debunked by the advent of novel Coronavirus also known as COVID-19 when the outbreak started from oriental region of China and then spread across occidental Europe before engulfing the entire globe. In such an unpredictable scenario, medical science would have to take aid of probabilistic modeling and analytical tools to gauge the risk involved and to evaluate

and estimate the potential trammel of the pandemic. Exceedance probability curves are one such utilitarian piece mechanism which has the capability to amalgamate the probability of a threat and its corresponding economical loss. Taking an example of one such potential threat, influenza, which has the capability to become a pandemic that can cause 6 million deaths worldwide, it was observed with the help of the EP analysis that in a given year, the probability of the influenza taking the form of a pandemic comes as 1%, which is a potent risk [2, 5, 6].

The healthcare analysis domain can also help in laying down the potential impact which a pandemic might cause in a stipulated period of time. It is observed that an outbreak in the range of a pandemic can cause colossal and outstretched growth in mobility and mortality. It has also been observed that in lower and middle income countries, this kind of an outbreak can have excessive mortality impact. There are a number of verticals and channels through which a pandemic can cause economic turbulence which includes short- and long-term setbacks. The short-term setbacks are usually caused due to the fiscal shock a market receives during the spark period. The long-term setbacks are the result of damage to the perennial economic growth during the spread phase and beyond that. Talking about the physiological impacts, a pandemic may induce behavioral changes which lead to widespread panic and agitation. They are widely interconnected to the long-term setback on the economic market as the fear results to a state of reluctance toward investments and purchase. In certain specific countries, a pandemic can also give rise to political contortions which in turn results to mob violence, mass clashes and a state of constant tension between the citizens and the government. In the view of all these medical, social, economic, physiological and political impacts, healthcare analytics can play a vital role as a study and chieftain combat mechanism by helping with the personalized evidence to enable pro-active decisions.

4 Development of Healthcare Information System and Healthcare Analytics

In order to understand the current implementation of healthcare analytics, it origin and consecutive journey needs to be traced. The example of the evolving healthcare information system has to be taken in order to understand the current state of healthcare informatics. The earliest record of a healthcare information system dates backs to 640 BC where the traces can be found in form of case studies. They portrayed the course of illness though observation of patients symptoms. In order to perform this practice, openness and honesty was required by the patient and observer. It was not a very successful practice and frequent deaths were encountered [7].

In modern times, the roots of information technology in healthcare sector were laid in 1960. For this study of the journey of information technology in healthcare sector, the rise of IT in the healthcare domain of the USA would be observed as model country for the world. The 60 marked the enactment of Medicare and Medicaid into practice.

In July 1965, President Lyndon Johnson signed the landmark bill which is prevalent till today and is referred as Medicare [8]. These two laws gave a strong ground for healthcare information system to emerge. Medicare is a federal program which is intended to provide healthcare coverage to US citizen with greater than sixty-five years of age or having disability regardless of the income. Similarly, Medicaid is state and federal program which provides health coverage for people with extremely low income. Courtesy to these two program, the practice of cost-based reimbursement started, and hence, the US government started keeping a lot of medical and peripheral records. In this decade, significant healthcare expansion also took place. There were a good amount of financial needs and capturing revenues specifically to the healthcare sector. The state of information technology was considerably primitive in this decade. The work was done on mainframe computers. Since mainframes were bulky in size, they were not penetrating through an environment brimming with manual records. Lack of portability and esoteric technical specification thwarted the use of mainframe computers in this decade. Central processing of data was a common practice of this decade which changed in the next ten years. Also, the presence of only a few vendor-based products also contributed in limiting the use of informatics technology services in the field of healthcare. There were presence of administrative and financial systems which mostly started because of the prevalence of Medicare and Medicaid. Talking about the infrastructure, there were presence of large hospitals and medical centers which required presence of information technology solutions. All the solutions which performed centralized data processing on mainframe computers were developed and maintained in-house [8, 9].

The 1970 marked a colossal growth in hospitals and medical centers and there were huge opportunity blooming for this sector. The rising Medicare and Medicaid expenditure exhorted a requirement for a digital solution of higher efficiency [9]. There was a growing need of cost containment in this sector which could only be achieved by replacing mainframes by a less bulky structure and a higher penetration rate among masses. This requirement was fulfilled by the advancement of minicomputers which were like small mainframe computers. Since the portability issue was solved, minicomputer became a lot popular. This decade also marked the dawn of computer network which meant that different computers present in hospitals, independent clinics and the government offices could now be connected and information could be shared instantaneously. This was the decade which also marked the availability of turnkey systems though vendor community. Turnkey were highly personalized computer systems which were customized to function for a specific application. They were mostly industry specific and healthcare industry had a surging demand for such systems. The term turnkey was gleaned from the objective that user can make the system ready by just turning a key. In contained, all the hardware and software required to function for a specific industry or requirement. Because of the presence of such highly personalized devices and solutions, there were an augmented interest in clinical applications. The boom of information system reduced the costs to an extent that it was affordable for medium-sized practitioners and hospitals to use and benefit from these solutions. The shared system which marked its dawn in this decade is prevalent and used till now [8, 9].

The 1980 brought the computers to the vicinity of masses and it had a significant impact on the healthcare industry as well. In this decade, the Medicare program introduced the diagnosis-related group which revolutionized the way information system was perceived in the healthcare industry. The diagnosis-related group was initially implemented across the USA by an organization called as Healthcare Financing Administration (HCFA) [9]. It was aimed at bolstering the cost controlling for inpatient services which was associated with and billed to Medicare. The organization Healthcare Financing Administration (HCFA) started to use DRGs after the development of the preferred plan for provider. This particular instance was started in 1989 and gradually was carried to other HMSA plans. A diagnosis-related group can be defined as a system whose motive is to classify hospital cases into various groups for better processing and storing of information. There were a total of 467 groups when this system began, and usually, the last group denoted cases which cannot be placed into any other groups. The initial motive behind developing this group-based system was to persuade the US congress to allow its use for reimbursement process in Medicare and Medicaid. Before this methodology was adopted, the reimbursement process was "cost-based" and DRGs were developed to make it "group-based" and hence add a certain layer of automation to it. This initiative was very successful and is used till now in determining how much the Medicare program would pay the hospital for carrying out a service on a patient. By dividing the cases into groups, a lot of time and resources were saved since the cares in each groups were clinically similar in nature. Also, same group cases tend to use the same level of hospital resources. Apart from DRGs, once major changed which was a harbinger of the state of information technology in healthcare sector was the emergence of personal computers. The unveiling of PCs brought a sea of change with itself. Networking became more sturdy and structured with the segregation of local area network (LAN) and wide area network (WAN). Later in the decade, the increased use of personal computers led to the decentralization of data processing. The billing system was also introduced which became more common in the next decade. All of the abovementioned development in the healthcare information system led to expansion of clinical information systems in hospitals [8, 9].

The 1990 was a turning point in terms of information technology across the globe and it had multifaceted effects on the healthcare sector as well. This decade saw a surging growth in the managed care sector and the integrated delivery systems. Also, another significant step taken by the Institute of Medicine (IOM) was the call for "computer-based" patient and the EMR [9]. The EMR or electronic medical record began as an idea or concept of recording and later storing patient data and information in electronic form, instead of the conventional method which involved a paper. Though the roots of this concept date back to 1972 when the Regenstreif Institute in Indianapolis developed this idea and labeled it as a huge advancement in healthcare and medical practices. Despite of being developed in the early 70, this concept was not used widely because of the associated high costs. This thwarted its mass use and limited the scope of operation to certain specific government hospitals [8]. Later, in 90, this concept became the backbone of healthcare analytics and increased its scope and usage. Also, 1990 marked the advent of the World Wide

Web. It was an era where efforts of certain IT giants led to considerable drop in cost of hardware. The proliferation of Internet has a positive impact on the healthcare industry. The use of Internet allowed the clinic and hospitals to skip the humongous costs of setting up a specific network for some of their intra-domain practices. The healthcare organizations took advantage of this Internet boom and a study by the Institute of Medicine predicted that by the end of this decade, there would be at least one computer at every physician's office which would be dedicated to use of computers for improving the patient care [8, 9].

The 2000 led to strong reforms which established both the necessity and presence of information system and analytics in healthcare industry. During this decade, the international organization gave a report on patient safety and medical error which demanded policies which could address the same issues. This decade marked a turmoil in the all of the major industries and healthcare industry was not spared, as the world faced spiraling health cost. It was also during this decade when the Technology Informatics Guiding Education Reform (TIGER) worked to advance the integration of informatics in the healthcare industry with the augmented application of information technology to improve the patient care while keeping an environment of learning health systems. Founded in the year 2004, the TIGER initiative was a revolutionary step in marking the goals and requirements toward allowing the various facets of the healthcare sector to use informatics tools, theories, principals and practices. These technologies were infused and interweaved together into pedagogy, research and practices which aimed at effective outcome of patient cases, their safely and cost reduction and optimization. In the USA, this was a time of economic upheaval and a growing number of citizens were uninsured. In order to mitigate the damage done by the economic recession of 2008, the 111th US congress signed a law called the American Recovery and Reinvestment Act of 2009 which also had provisions to boost up the healthcare sector [10]. As part of the American Recovery and Reinvestment Act, another program was enacted, namely the Health Information Technology for Economic and Clinical Health (HITECH) Act. This act ensured that proper measures are being taken to promote the adoption of information technology into the healthcare sector. In the following year, the landmark bill of the Affordable Care Act, formally known as Patient Protection and Affordable Care Act, and popularly called as the Obamacare was signed by the 111th US congress [9, 11]. This was consecutively signed into law by then President Barack Obama and was part of his vision to transform the healthcare sector. The various provisions present as part of Obamacare ensured that it allowed surplus funding to promote information technology into healthcare domain which resulted into the structured storing of clinical and medical data [8–11].

The last decade has witnessed a massive digital growth and has engendered something we call as the big data. While big data can be defined as the structured and non-structured set of enormous volume of data, "big data in healthcare" refers to the colossal health data amassed from innumerable wellsprings or sources. The data includes electronic health records (EHRs), genomic sequencing medical imaging, wearables, payer records, pharmaceutical research, medical equipment and so on. There are three specific characteristics which differentiate this segment of data from

conventional electronic medical and human health data used for decision-making. This first characteristic is the availability in remarkably huge volume. The second characteristic is that it moves at high velocity and covers the healthcare sector's mammoth digital universe. The final characteristic lies in its diversity. Since it gets fetched from innumerable sources, it is highly variable in structure and nature. The above three characteristics are known as the 3Vs of big data in healthcare sector. Since it brings with it a profound diversity in format, type and context, it is onerous to merge big healthcare data into conventional databases such as the ones used in the previous generations to store medical and clinical records which were DB2 and IMS DB. This makes the task to process this data very challenging. Also, industry leaders find it arduous to utilize or mobilize its considerable potential to transform the healthcare industry. Despite these challenges, there are several new technological means and methodologies which allow the conversion of big data in healthcare sector utilizable and actionable commodity [12].

Once the appropriate set of big data is secured, the next steps involved its convergence which make it ready to be processed and used as a solution [12]. If one has to define the process of convergence in a hackneyed terminology, it can considered as the amalgamation of two very different entities. In the backdrop of analytics, it is the integration of two or more than two distinct technologies, commodities, data or systems into a single unit, device or system. One example of this process is emergence of a cellphone with a camera functionality. The process of convergence here has amalgamated a traditional camera and communication device to create a new piece technology or a device. One prime example of the use of convergence is the one currently take into task by Deloitte where one of its particular divisions is working on the convergence of healthcare trends [13]. This kind of convergence is transforming the traditional US healthcare industry using four major convergence trends. This is creating opportunities for innovation in the healthcare analytics sector. The four major trends are everywhere care; wellness and preventive care; aging, chronic and end of life care and personalized care. The convergence methodology is being used in the everywhere care trend to transpose the gamut of care from hospitals to comparatively lower-cost sites. The next trend wellness and preventive care focuses on repositioning the disease management from reactive to preventive. The third trends focused on utilizing big data to personalize and hence manage chronic conditions to provide support in aging, chronic and end-of-life care. The final trend of personalized care focuses on transmuting the healthcare services from mass generalization to mass customization which would add an extra layer of precision to it. A lot of organizations are considering to develop similar innovation strategies across the healthcare domain to remain afloat in this ever-changing scenario and landscape, by capitalizing on the emanating opportunities [13].

5   Results

After passing through the convergence technique, various algorithms can be applied onto the datasets to garner insights about the information, which a certain dataset wants to convey. For the datasets related to pandemics, a lot of information can be gathered by applying these algorithms on the cleansed and structured data which would provide trends and information about the worst-affected areas, countries or continents. They would also shed a light on the mortality rate and its metastasize across different places. The information could be generated by using a single piece of technology for data analysis or making histograms such as R or Python. To check the information provided by these histograms, data analysis was performed on several sample datasets of early twentieth century's Spanish Flu and COVID-19. The following set of graphs were developed using Seaborn library which is a tool for making statistical histograms in python. This library builds on top of matplotlib and consolidates intimately with pandas data structure. The plotting is carried out by firstly performing the semantic mapping and statistical aggregation on the dataframes and arrays which contain the whole dataset. This process results in the formation of informative plots. This technology rather than focusing on the details on how to draw a plot, prioritizes on defining what the different parts of the plot refer to. This technology is dataset-oriented can be considered as a declarative application programming interface [13, 14] (Fig. 1).

The above set of statements import the necessary libraries required to perform the analysis on the given dataset. Once the importing is done, the visuals would be created using another set of statements (Figs. 2 and 3) [14].

The above set of statement set the columns representing districts, dates and mortality as pivots. The dataset represents district-wise information about the mortality during the Spanish Flu phase. The data indicates the number of deaths per week due to fever in the districts of Assam, India, from 1916 to 1921. The numbers

Fig. 1 Code snippet of the import statements

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
```

Fig. 2 Code snippet of setting the pivots

```
[ ]  df = pd.read_csv('sp_1.tab', delimiter='\t')
     df.Date = pd.to_datetime(df['Date'])

     df = df[['District', 'Date', 'Mortality']]

     df['Year'] = df.Date.apply(lambda x : x.year)
     df['month'] = df.Date.apply(lambda x : x.month)
     df['day'] = df.Date.apply(lambda x : x.day)
```
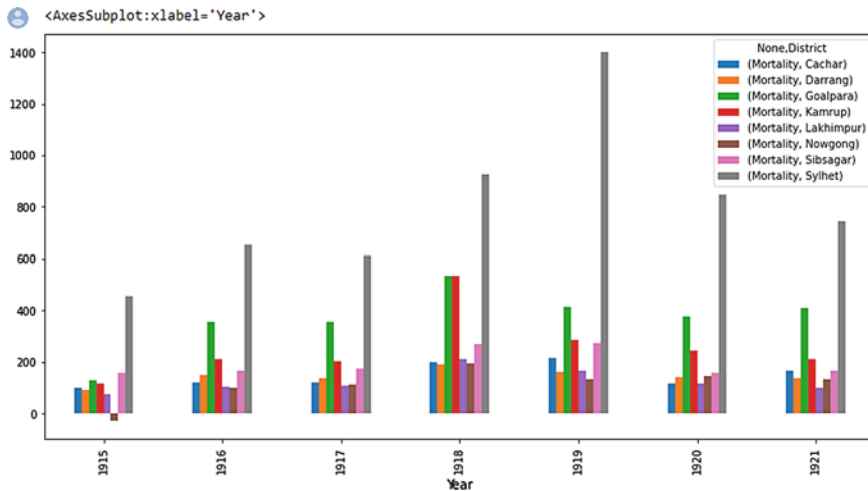
Fig. 3 Plot for average mortality for certain districts

are seasonally adjusted to show excess mortality due to the influenza pandemic of 1918–19 [15, 16].

The data here is a multi-level group by in order of Year and District to calculate the average mortality for those levels. The visual created is done by unstacking the pivots created before. This is done to convert the inner most labels as new column labels in the dataset. After that we have on the x-axis the years with each year representing the average mortality for each district. The following is not so different from a bar graph except that it contains multi-label for each first-level attribute upon which the data is grouped.

The above plot represents the average mortality of the districts of Cachar, Darrang, Goalpara, Kamrup, Lakhmipur, Nowgong, Sibagar and Sylhet from 1915 to 1921 during the trepidation period of Spanish Flu (Fig. 4) [16].

The above plot represents the month-wise mortality of the districts of Cachar, Darrang, Goalpara, Kamrup, Lakhmipur, Nowgong, Sibagar and Sylhet for a calendar year (Fig. 5) [16].

The above plot represents another view for the month-wise mortality of the districts of Cachar, Darrang, Goalpara, Kamrup, Lakhmipur, Nowgong, Sibagar and Sylhet for a calendar year.

Next, the data is separated for each mortality by each district and stored separately. These datasets are individually used to make uni-variate plots. The following plots are modeled upon the Gaussian distribution to express the normalness of the data. Each of the district's mortality is plotted in the same pyplot figure to give a comparative understanding of the mean and skewness of the rates in each district (Figs. 6, 7, 8 and 9) [16].

The next set of operations were performed on a dataset of COVID-19, where the total cases, recoveries and deaths were delineated across the top ten worst-affected
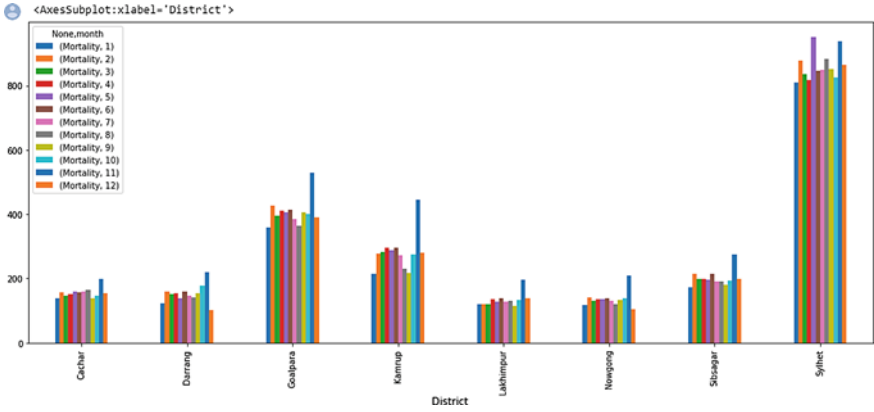
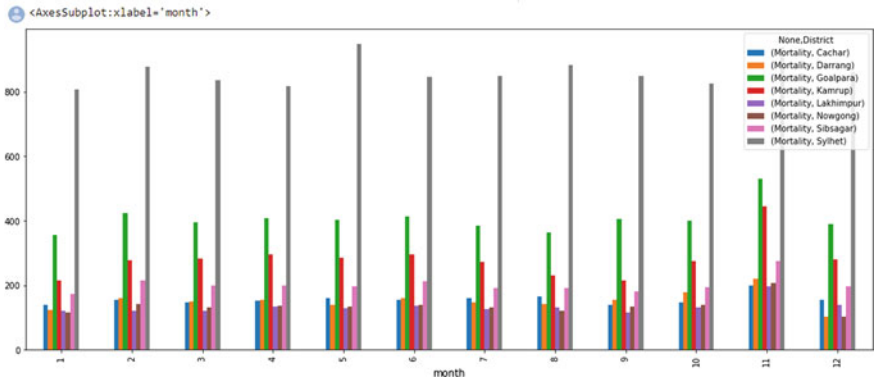**Fig. 4** Plot for month wise mortality for certain districts. Dataset: Mortality during the Spanish Flu

Fig. 5 Plot for month wise mortality for certain districts. Dataset: Mortality during the Spanish Flu
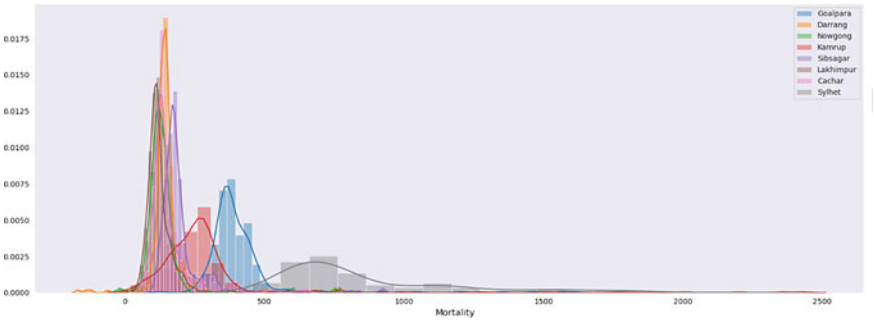

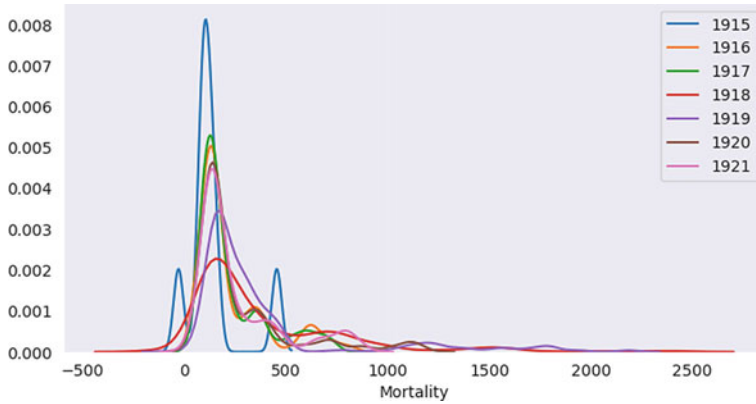**Fig. 6** Distribution of mortality across districts. Dataset: mortality during the Spanish Flu

Fig. 7 Distribution of mortality across years. Dataset: mortality during the Spanish Flu
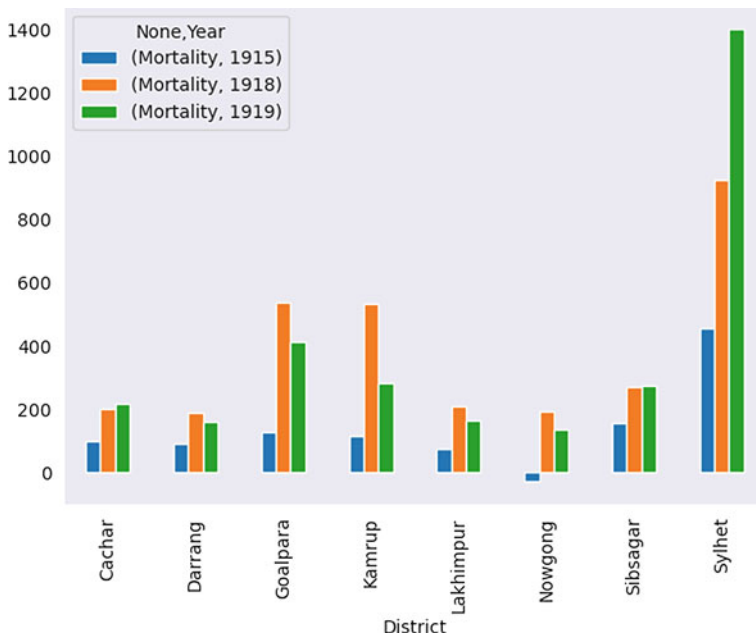

Fig. 8 Bar graph representing mortality across three specific years. Dataset: mortality during the Spanish Flu

countries. The dataset used is the COVID-19 cases, deaths and recovery repository from Worldometer. The dataset itself has been captured from a number of sources and also doubles up as a leaderboard to provide an updated count of the cases, recoveries and other information about COVID-19. The list of countries present in the dataset
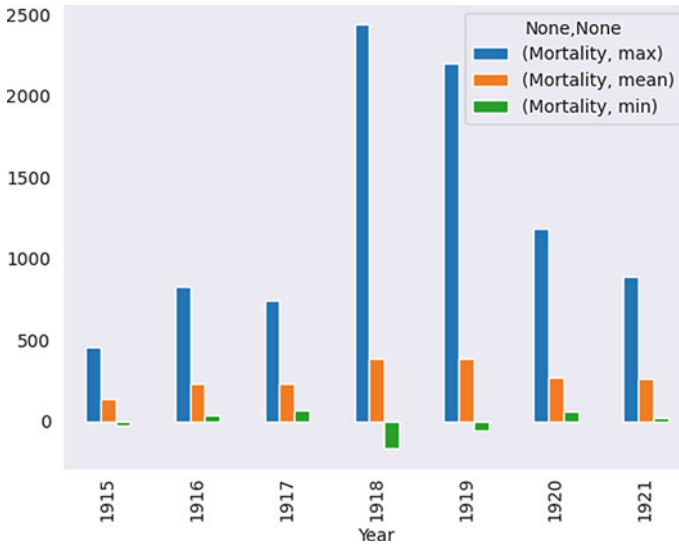
Fig. 9 Bar graph representing the max, min and mean mortality from 1915 to 1921. Dataset: mortality during the Spanish Flu

and territorial boundaries including continental regional classification has been done in accordance to the United Nations Geoscheme [17] (Figs. 10 and 11).

Once all the necessary data is accumulated, converged and processed, it has to be converted to an applicative software solution which contains multidimensional functionalities to support the operational performance, clinical outcomes and optimizing overall efficiency of healthcare as a service. A lot of operational and business factors need to be taken into account for developing a solution which gets a wide range of insights such as hospital management, costs, diagnosis and patient records. The majority of the healthcare big data has to be scrutinized and handled with utmost care in the most discreet way which would respect the patient's privacy and shield the sensitive information. During the case of a pandemic, all of these actions have to be carried out in a very short period of time, and hence, a proper business suite is required for the healthcare analytics solution to fight a pandemic. There are a number of ways in which a business intelligence software can transform the way a healthcare analytics solution is developed and used. These difference methodologies include financial planning, evaluating performance, taking care of a patient's satisfaction, coordinating communication, managing reputation, predicting the future, data visualization supporting and improving decision-making and so on. The solution should be able to manage the healthcare outcome in such a fashion that it get optimized to achieve the optimum use of analytics (Fig. 12) [15].

The above diagram explicates the optimum use of analytics as the amalgamation of the clinical outcomes and operational outcome. Unless these two sectors function together as an entity, the optimized use of analytics cannot be achieved. The first facet is the clinical outcome and there are a number of factors which have to be taken into
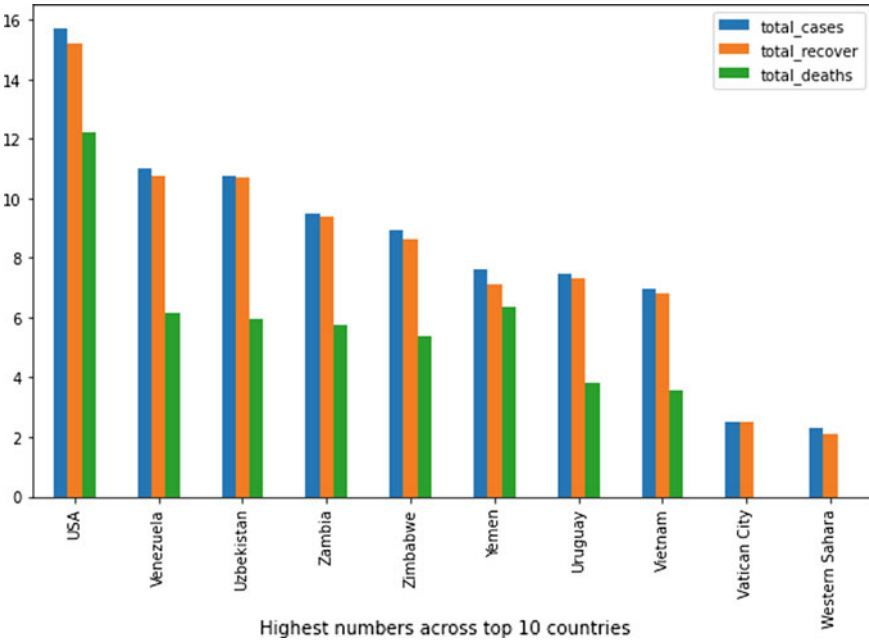
Fig. 10 Bar graph representing total cases, recoveries and deaths. Dataset: COVID-19 Worldometer
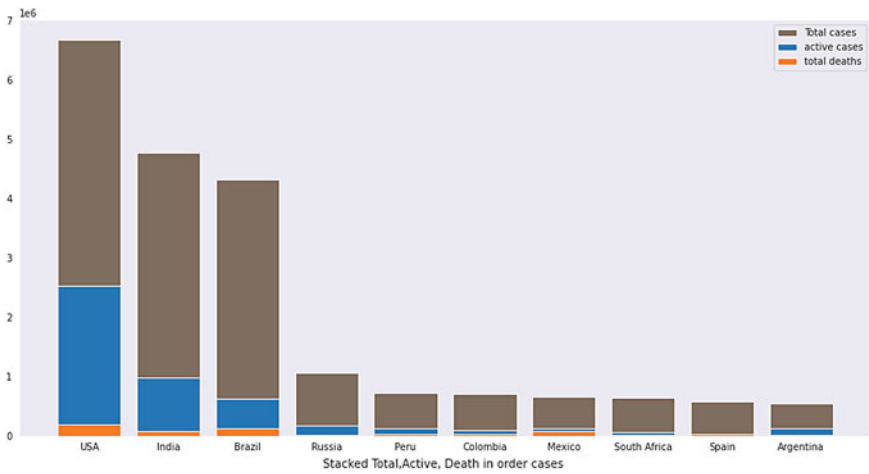

Fig. 11 Bar graph representing stacked total of cases, recoveries and deaths. Dataset: COVID-19 Worldometer

Diagnostic Assistance

Clinical treatment
effectiveness

Critical Care
intervention

Research for improved
Disease Management

**Clinical
Outcomes**

Operational
Outcomes

Readmission prevention

Claims Management

Fraud Detection and
Prevention

Voice of the patient

Patient Discharge and
Follow-up care

Optimum Use of Analytics

Fig. 12 Optimum use of analytics

consideration. The first facet is the diagnostic assistance. Analytical solution should be able to help into providing a diagnostic assistance for any aliment or symptoms which generally appear during an outbreak of a size of epidemic or pandemic. Second comes the clinical treatment effectiveness which can be monitored using this solution. The solution can also provide a grading mechanism using which the effectiveness of various clinical treatment can be evaluated and compared. The next factor on the side of clinical outcomes is the critical care intervention. The optimized use of analytics would help in predicting the requirement for a critical care among innumerable cases. The use of an intelligent analytics solution can also help in the research for improved disease management. The operational outcomes also lie on a number of factors which need to be taken care of before declaring the analytics solution as intelligent or optimized. The analytical solution can help in the prevention of readmission of cases. In case of a pandemic, the worst hit areas get multiple waves and a lot of readmissions which need to be mitigated. Next is the claims management, which is something that analytical solutions have been taking care of since a long time in term of providing multiple DRG, Medicare or Medicaid solutions and so on. The solution should also be able to detect the cases of frauds and incorrect entries. The solution can also contain a feedback mechanism to be the voice of patients which can provide trends on which hospitals or clinic is performing better than the rest. At last, the solution should look over the patient discharge and follow-up care [14, 18].

6 Illustration

In order to achieve the business transformation as mentioned in the previous section, a business solution must adopt an end-to-end approach toward analytics and not just follow one aspect. This end-to-end approach can be achieved by incorporating four specific layers which are business context and planning, analytics modeling, data layer, technology layer. The below diagram shows sample architectural diagram of a
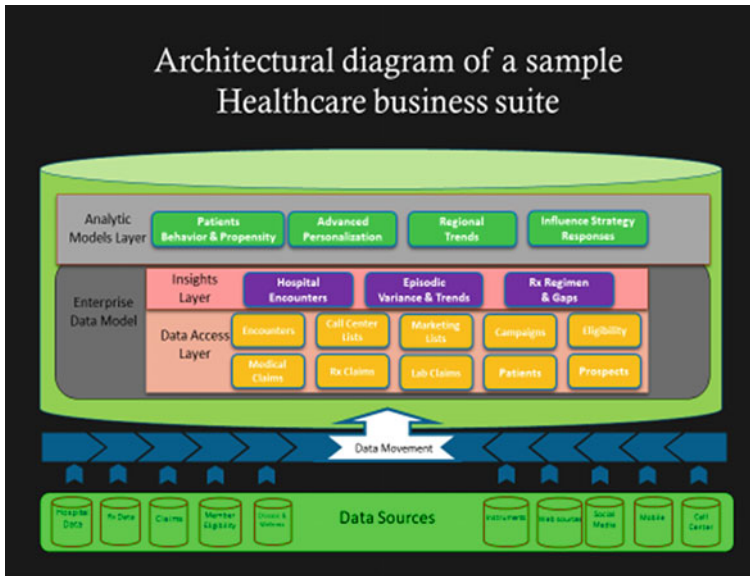
Fig. 13 Architectural diagram of a sample healthcare business suite

healthcare business suite which contains the required layer for fulfilling the business transformation [19] (Fig. 13).

The raw data as we see in the diagram comes from various sources such as hospitals, prescriptions, claims, member eligibility repository, disease and wellness repository, medical instrument records, web sources, social media, mobile, call centers and so on. In the diagram, the medical prescriptions have been denoted by Rx claims. The symbol "Rx" found its origin form a Latin word "recipe" which meant "to take." The Rx claims data is mostly formed by the customary part of a superscription which is the heading of a prescription. Next is the enterprise data model layer which has been further segregated into two parts—data access layer and insights layer. The enterprise data model layer contains an amalgamated view of both the data provided and used across the organization. This layer is responsible for incorporating the business standards, SOPs and an appropriate perspective of the healthcare industry. It also responsible for representing an unbiased view of a single integrated definition of data. This layer formulizes and brings together all the entities which are important to the sector or organization and the rules which govern them. This layer does not taken into account how a single unit of data or a chunk of data is physically stored, fetched, accessed or processed. The framework of these layers revolves around integration. It facilitates the identification of data within and outside the organizational boundaries which are sharable and the ones which are redundant. While looking upon all the layers in tandem, this layer can be label as the starting point of all system designs. It is more of a blueprint which contains maps of all the processes and provide a whole

visualization of planning, development and implementation of the analytical solution. In the above sample architectural diagram, it is evident that all the raw data such as encounters, call center list, marketing list, campaigns, eligibility, medical claims, Rx or prescription claims, lab claims, patients and prospects have been clubbed together in the data access layer. While the more processed information such as hospital encounters as a whole entity, episodic variance and trends, Rx regimen and gaps have been clubbed together in the insights layer. The next layer is the analytic model layer which is a key element and is indispensable in comprehending the business data. This layer helps in making precise data-based predictions, and to extract the high-end insights which in turn, helps in making the correct business decisions. In a scenario where an analytical business solution is responsible for combating a pandemic, all of its internal layers should work in sync with each other to provide an optimized solution [14, 18, 19].

7   Conclusion

The methodical use of healthcare analytics has the potential to prepare mankind for the next pandemic. It also has the capacity to fight the current pandemic and become an effective tool of both defense and resurrection. However, the crux of the usage of this approach depends on a number of factors. It is essential for all those factors to be aligned in the correct form for this proposition to provide optimum results. To start with, organization and government policies should be favoring the approach of healthcare analytics. There should be adequate laws to facilitate and envisage the surfeit funding for the health information systems. Next, the raw data should be precise, voluminous and diversified to take in account a majority of the world population facing the pandemic. The gigantic amount of data has to be converged to be used as an input methodology for any analytical business suite. The data would have to be processed in accordance with the needs delineated in the enterprise data model of the solution. At last, a model solution keeps all the facets in sync which are essential in providing the optimum result. If the previous mentioned steps are carried out correctly, then the healthcare analytics software solution can fulfill applications such as diagnosis, preventive medicine, precision medicine, medical research, reduction of adverse medication events, cost reduction, population health monitoring and so on and so forth. The diagnosis process can be achieved by employing the means of data mining and analysis to identify the cause of illness. Significant success can be achieved in the course of discovering or detecting a preventive medicine for pandemics by using predict analytics and data analysis of lifestyle, genetic and social situations and circumstances to avert an outbreak. The aggregate data can be leveraged to drive a hyper-personalized care and provides precision medicine to masses. Healthcare analytics can bolster the medical research by employing a methodology of data-driven medical and pharmacological research to cure diseases and unearth novel treatments and medicines. The healthcare analytics can also contribute in the reduction of adverse medication events as the present of big data repositories can

be channelized to find medication erratum and highlight potential adverse reactions. Identification of courses, actions and methodology which has the potential to be a cost effect optimized solution can drive toward long-term savings.

The statement that big data is transforming the healthcare industry is indisputable to the extent that the statement itself is almost considered canonical. But, similar to any other technology, healthcare analytics has also got its fair share of limitations. The first limitations comes in the form of weighted average approach which a lot of data models follow. Now, there are a number of sectors such insurance sector which rely on the actuarial models for risk-management. The mechanism used in the healthcare analytics system can only refine actuarial models till a certain juncture. Since a sizable amount of healthcare and clinical data is unstructured, we a get a resulting dataset which is not normally distributed, and hence, the weighted averages concept cannot be used in the resulting models. Instead, one needs to find datasets for different subgroups and in order to look at a minuscule level. It is a quite regrettable fact that till today, a number of big data extraction tools are not adept enough to analyze data on such a granular level which makes this scenario a limitation of healthcare analytics. The second limitation comes while predicting or accessing a doctor's performance using the patient data. A lot of doctors refer a high-risk or vulnerable patient to their colleagues to improve their track record. This practice hampers the accuracy of assessing a doctor's performance since in this case, the records have been compromised by the human action. This is also a limitation of healthcare analytics as we do not have a riveted mechanism to mitigate or appease this action [12, 14].

Despite all these limitations, the healthcare analytics domain has got the potential to monitor the health of masses to identify disease trends and can provided enhanced health strategies based on demographics, geography and socio-economics. Hence, healthcare analytics is indeed an advent to counter the imminent threats of outbreaks such as epidemics and pandemics, provided it gets the necessary path of action to fulfill its grail.

References

1. 20 of the worst epidemics and pandemics in history. Retrieved from: https://www.livescience.com/worst-epidemics-and-pandemics-in-history.html.
2. Gori, D., E. Boetto, and M. P. Fantini. 2020. The early scientific literature response to the novel Coronavirus outbreak: who published what? https://doi.org/10.1101/2020.03.25.20043315v1
3. Kagan, D., J. Moran-Gilad, and M. Fire. 2020. Scientometric trends for coronaviruses and other emerging viral infections. https://doi.org/10.1101/2020.03.17.995795v2
4. WHO. 2020. WHO Timeline-COVID-19. https://www.who.int/news-room/detail/27-04-2020-who-timeline—covid-19.
5. Centers for Disease Control and Prevention (CDC). 2009. Swine influenza A (H1N1) infection in two children–Southern California, March-April 2009. *Morbidity and Mortality Weekly Report* 58 (15): 400–402.
6. Centers for Disease Control and Prevention (CDC). 2010. The 2009 H1N1 Pandemic: Summary Highlights, April 2009–April 2010. https://www.cdc.gov/h1n1flu/cdcresponse.htm .

7. Andreasen, V., C. Viboud, and L. Simonsen. 2008. Epidemiologic characterization of the 1918 influenza pandemic summer wave in Copenhagen: Implications for pandemic control strategies. *Journal of Infectious Diseases* 197: 270–278.
8. Griffin, J. 2017. The History of Medicine and Organized Healthcare in America. Retrieved from: https://www.griffinbenefits.com/blog/history-of-healthcare.
9. Burnham, J. C. 2016. Health Care in America: A History. Journal of American History, Vol. 103, Issue 1, June 2016, pp. 167–168. https://doi.org/10.1093/jahist/jaw175.
10. The Pros and Cons of Obamacare. Retrieved from: https://www.healthline.com/health/consumer-healthcare-guide/pros-and-cons-obamacare.
11. Oberlander, J. 2013. The Future of Obamacare. Obstetrical & Gynecological Survey: April 2013, Vol. 68, Issue 4, pp. 265–266.
12. Big Data: The Management Revolution. 2012. Harvard Business Review. Retrieved from: https://wiki.uib.no/info310/images/4/4c/McAfeeBrynjolfsson2012-BigData-TheManagementRevolution-HBR.pdf.
13. Convergence of Health Care Trends. Deloitte US. Retried From: https://www2.deloitte.com/us/en/pages/life-sciences-and-health-care/articles/convergence-health-care-trends.html.
14. An introduction to seaborn. Retried from: https://seaborn.pydata.org/introduction.html.
15. Koh, H.C., G. Tan. 2011. Data mining applications in healthcare. Journal Healthcare Information Management pp. 65–73.
16. Chandra, Siddharth. 2015. Weekly mortality data for Assam, India from 1916 to 1921, Harvard Dataverse, Retrieved from: https://doi.org/10.7910/DVN/29045.
17. COVID-19 Coronavirus Pandemic. Retried from: https://www.worldometers.info/coronavirus/.
18. Janke, A.T., D. L. Overbeek, K. E. Kocher, P. D. Levy. 2016. Exploring the potential of predictive analytics and big data in emergency care. Ann. Emerg. Med. pp. 227–236.
19. Andreu-Perez, J., C.C Poon, R. D. Merrifield, S.T. Wong, G.Z. Yang. 2015. Big data for health. *IEEE Journal Biomedical Health Information*, pp. 1193–1208.