

Big Data in Healthcare: COVID-19, EHRs, Uses and Challenges

Sai Harshita Yangam, Dr. B. Lakshmi*

Assistant Professor, National Institute of Pharmaceutical Education And Research (NIPER), NH 9, Kukatpally Industrial Estate, Balanagar, Hyderabad, Telangana 500037, India

DOI: [10.36347/sjebm.2020.v07i12.001](https://doi.org/10.36347/sjebm.2020.v07i12.001)

| Received: 19.11.2020 | Accepted: 02.12.2020 | Published: 07.12.2020

*Corresponding author: Dr. B. Lakshmi

Abstract

Review Article

COVID-19 pandemic is an indicator for healthcare industry of opportunities and challenges present with data and information. Big data is massive amount of information generated and gathered by modern technology characterized by 4V's (Volume, Velocity, Variety and Veracity). In healthcare, electronic health records (EHRs) generate massive amounts of data regarding a patient. Apart from EHR, -omics data and IOT are other sources for information regarding an individual's health. When these all sources are put together it generates big data. Countries worldwide are adapting to EHR and Information and Communication Technology (ICT) systems due to its immense ability in research, clinical trials, drug discovery and repurposing, public health and personalized or targeted treatment for a cost-effective and evidence-based patient care. With big data and AI or ML technologies, the process and management is greatly simplified evident in COVID-19 situation. Though big data has many advantages and opportunities associated with it, challenges in terms of data structure, storage, standardization, its governance and security or privacy are yet to be addressed.

Keywords: Big Data, COVID-19, Machine Learning, AI.

Copyright © 2020 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY-NC 4.0) which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use provided the original author and source are credited.

Big Data Usage around the World during COVID-19 Pandemic

The daily news of havoc brought by COVID-19 pandemic around the world is serious and worrying. To combat this global health emergency, information is vital for better decision making and organization. Big data lies at the core to understand and predict the impact the Coronavirus will have on all of us.

McKinsey reports on the essential part the chief information officers play in reaction to COVID-19 in different organizations. Businesses at the bleeding edge of this widespread are depending on information in numerous ways to reply viably. For occurrence, healthcare suppliers, are leveraging information from nations influenced early by COVID-19 to estimate needs for veils, clinic beds, and ventilators. Basic supply chains are utilizing deals information to help wholesalers distinguish and dispatch fundamental things to their clients. Telecom players are utilizing organize activity information to choose where to overhaul capacity in reaction to the incredible request for transmission capacity from an progressively inaccessible workforce. Information from Worldwide Situating framework (GPS) considers populace development by locale, city, etc., which makes a

difference in giving the population's compliance with social-distancing commands (Kevin Buehler).

Taiwan's reaction to COVID-19 included investigation of huge information by coordination its national wellbeing protections database with its movement and traditions database, this produced real-time cautions amid a clinical visit based on travel history and clinical side effects that helped in case recognizable proof. It utilized QR code filtering and online detailing of travel history and wellbeing indications to classify travelers' irresistible dangers. Persons with low risk were sent an SMS (short message service) for faster immigration clearance and those with higher risk were quarantined at home and tracked through their mobile phone to ensure they remained at home during the incubation period [1].

Canadian startup BlueDot is using AI to track infectious disease risk by analyzing data from news reports, social media platforms, and government documents. It had already warned of the threat several days before the Centers for Disease Control and Prevention or the World Health Organization issued their public warnings.

Google's DeepMind and BenevolentAI are using AI and computing power to develop treatments and to build drugs. Inside weeks of the flare-up, it utilized its prescient capabilities to propose existing drugs that could be valuable (Marr).

Chinese government created Health Code, a checking framework that employs huge information to distinguish and survey the chance of each person based on travel history, time went through in infection hotspots, and potential presentation to individuals carrying the infection. They were color-coded to show in case they ought to be isolated or not.

Tech companies like Tencent, DiDi, and Huawei have allowed access to their cloud computing resources and supercomputers to researchers to fast-track the development of cure or vaccine.

INTRODUCTION TO BIG DATA

Data is key for superior organization and unused improvements. The more information, the more optimally we organize ourselves for better outcomes. Data collection is very crucial for any organization [2].

Massive amount of information is generated and gathered by modern technology. Data in vast quantities is collected through social media, smartphones, innovative sensors, DNA analysis, chemical screening, data from genome and proteome,

climate data, screening data for promising drug candidates, data on health, both of a person and the population [3].

This driven to the creation of the term "big data" in healthcare industry that's huge and unmanageable. To meet our display and future needs, we ought to create unused methodologies to organize this information and determine significant data.

Need for Big Data

International Data Corporation (IDC) estimated the approximate size of the digital universe to be 130 Exabytes (EB) in 2005. This digital universe expanded to about 16,000 EB or 16 Zettabytes (ZB) in 2017. IDC predicted that the digital universe would expand to 40,000 EB (40 ZB or 40 trillion GB) by 2020. This can be envisioned by assigning about 5200 Gigabytes (GB) of data to all individuals. This reflects the pace at which the digital universe is expanding [2].

Defining Big Data

Gartner (circa 2001) [5] defines Big data as, "Big data is high-volume, high-velocity and/or high-variety information sets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation". Hence, 3V's were initially considered as characteristics of big data. Later veracity, value and variability were added.



Fig-1: 6V's of Big Data

IBM defines big data in terms of 4V's to achieve superior value from data analysis. These include volume, velocity, variety, and veracity. Veracity

refers to authenticity and reliability. Later, the value was also included. Data has intrinsic value. But it is of no use until that value is discovered.

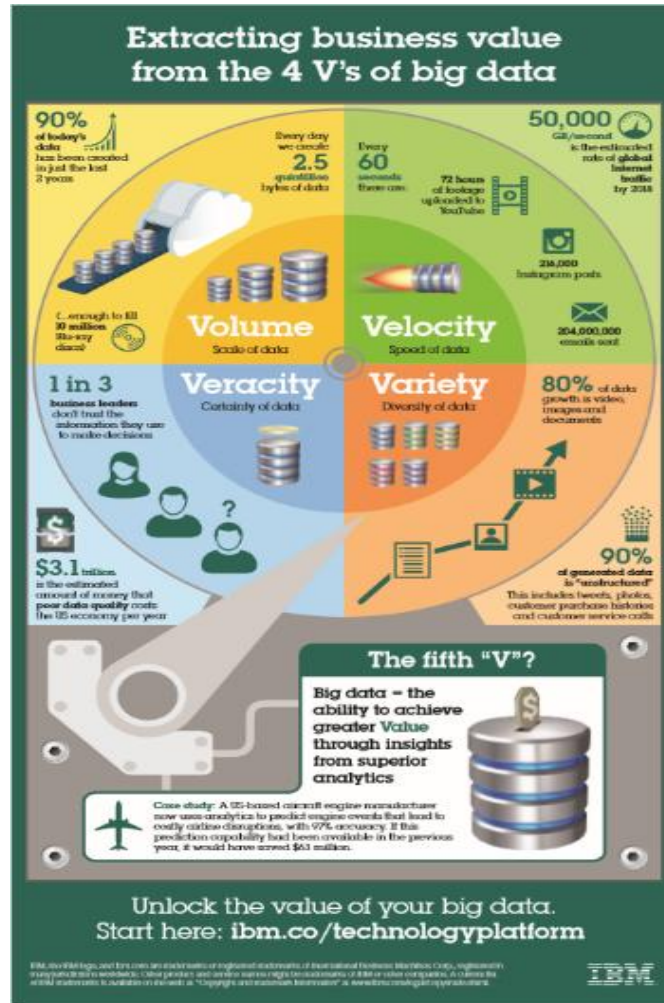


Fig-2: IBM - 4V's of Big Data

Characteristics of Big Data

Table-1: Characteristics of Big Data [4]

CHARACTERISTIC	DESCRIPTION
Volume	The amount of data matters. With big data, you will have to process high volumes of low-density, unstructured data. This can be data of unknown value, such as Twitter data feeds, clickstreams on a webpage or mobile app, or sensor-enabled equipment. For some organizations, this might be tens of terabytes of data while for others it may be hundreds of petabytes.
Velocity	Velocity refers to the rate at which data is received and perhaps, acted on. Normally, the highest velocity of data streams directly into memory versus being written to disk. Some internet-enabled smart products operate in real time or near real time and will require real-time evaluation and action.
Variety	Variety refers to the many types of data that are available. Structured data – database, data warehouses, CRM, ERP etc. Unstructured data – images, video, presentation, mp3 files etc. Semi-structured data – XML, CSV, JSON etc.
Veracity	Veracity refers to the data authenticity and reliability. Ambiguity within data is the primary focus in this dimension – typically from noise and abnormalities within the data.
Value	Refers to the value the data holds
Variability	Consistency of data over time

Different sectors using Big Data

In today's competitive environment. From retail to healthcare, real estate to agriculture, and finance to telecom, sports to healthcare, and energy to utilities there are several industries harnessing the power of big data to create value from increasing flood of data generated.

With tons of industries using big data to convert the growing influx of data into useful insights and gain competitive advantage in the bug data evolution – big data is making waves in almost every sector.

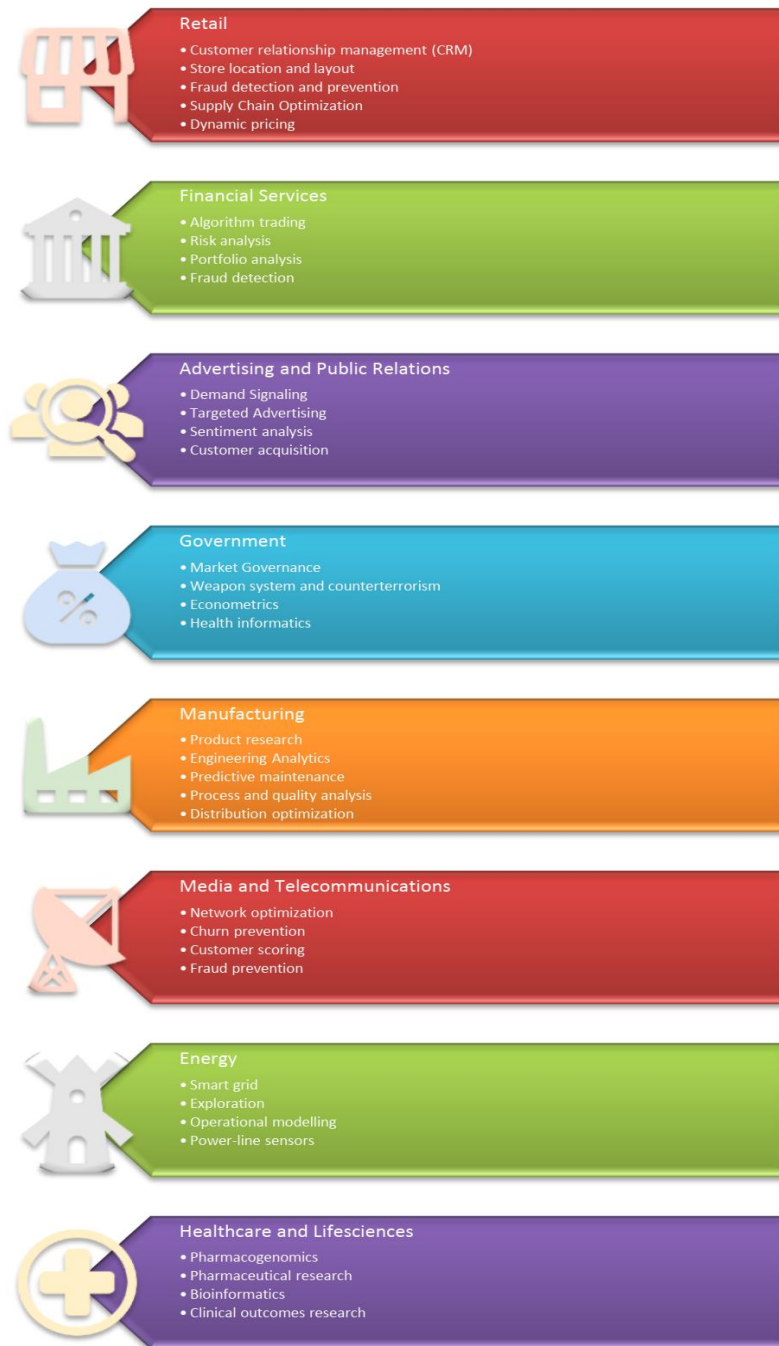


Fig-3: Various Industries using Big Data

OBJECTIVES OF THE STUDY

- To understand the role of Electronic Health Records (EHRs)
- To understand how big data is analyzed and managed
- To identify its uses along with COVID-19 as an evidence
- To identify the challenges, it poses

BIG DATA IN HEALTHCARE

In healthcare, with the availability of big data and the ability of supercomputers, it enables users or researchers to advance in terms of predicting epidemics and pandemics, its spread (such as COVID-19), effectiveness and side effects of novel drugs, developing personalized treatments (based on patient history and big data), reducing treatment cost, integration of machine learning methods and other AI techniques to mention few. In spite of various preferences, there too exist challenges in terms of execution. For illustration, social and lawful challenges in terms of security, information possession, administration etc [6].

Electronic Health Records (EHRs)

A National Institute of Health report in 2006 defines EHR as a longitudinal electronic record of quiet wellbeing data produced by one or more experiences in any care conveyance setting. It incorporates socioeconomic, advance, issues, drugs, imperative signs, therapeutic history, research facility information and radiology reports of a persistent. Basically, EHR could be a advanced form of a patient's wellbeing record.

Medical records were present earlier since the advent of healthcare. It was used to record the disease and its probable cause. In US, during 1960's and 1970's healthcare industry was undergoing a rapid change when the federal government passed law for establishment of Medicare. During this time, healthcare market was improving, lawsuits started emerging, quality became significant and government was stringent with regulating the healthcare industry. Medical records became necessary but its adaptation by healthcare providers and physicians was slow [7]. When American Recovery and Investment Act of 2009 was passed, it offered incentives to all healthcare providers and facilities for implementation of Electronic Health Records (EHR). In 2014, 76% of healing centers embraced EHR frameworks and 97% intense care clinics have received certified EHR framework. The "All of Us" extend is extending upon UK biobank demonstrate by making a coordinate interface between persistent genome information and their phenotypes by joining EHRs, behavioral, and family information into a special persistent profile [8].

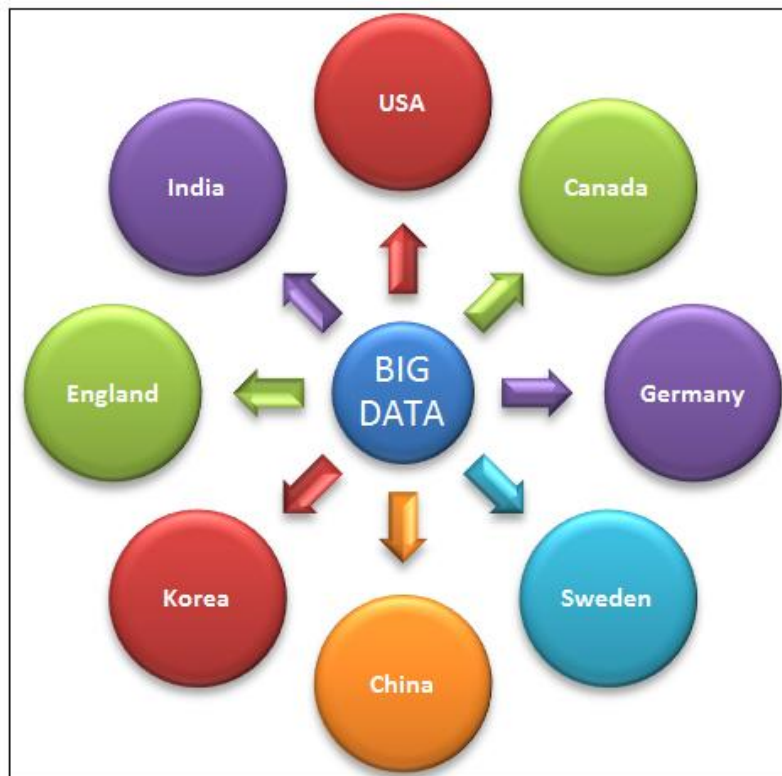


Fig-4: Countries adopting to EHR System

In 2001, Canada modernized its Information and Communication Technology (ICT) infrastructure in healthcare. 91% Canadians had EHRs and 91,000 clinicians were using EHR systems by 2015 [9].

The FAIRplus extend was kicked-off to progress information administration and information administration in European wellbeing investigate. It points to extend the disclosure, availability, and reusability of information from chosen ventures supported by the EU's Innovative Medicine Initiative

(IMI), and inside information from pharmaceutical industry accomplices. Janssen Pharmaceutical NV is the pharmaceutical industry lead of the extend. The venture includes 22 pharmaceutical, little company and scholastic accomplices from nine European nations. It is supported by the IMI as a public-private organization and will run for three and a half a long time (2019-2022).

Britain with its “National Plan for IT” in 2001 modernized healthcare framework. Beneath this, Summary Health Records (SCRs) were made for 54 million, about 96% of the populace. The UK Biobank extend has collected wealthy information on 500,000 people, collating together organic tests, physical measures of understanding wellbeing, and sociological data such as way of life and socioeconomics. It propelled 100,000 Genomes venture with the aim to get it the hereditary beginnings being common cancers [8].

In 2016, about 90% of physicians started using EHR systems in Germany. Patient Privacy was made significant where a patient can decide whether to hide or block any entry in the health record [9].

Most of the tertiary hospitals in Korea are using EHRs. A nationwide Health Information Exchange (HIE) platform is built for exchange of information among healthcare facilities.

Sweden and China are moreover embracing this model—leveraging one of a kind character numbers issued to citizens to connect something else detached datasets from authoritative and healthcare records. In this way, tests, innovations, and strategies will be coordinates in a way that's particular to the quiet but not fundamentally to the clinic or clinic.

Mixed healthcare system consisting of hospitals run by Central Government, State Government as well as Private, exist in India. The level

of use of ICT in the country has been lower compared to other countries. Corporate hospitals in India such as Max Health, Apollo, Fortis, Sankar Nethralaya, and many other have implemented ICT systems in place, covering all aspects, i.e., registration, billing, laboratory and clinical data [9].

In US and other created nations, national and state registries collect wellbeing related information and total it over a long time, hence making enormous information accessible for researcher to work on. In India, wellbeing records are amassed and put away by person wellbeing organizations subsequently there's a plausibility of duplication and trouble in get to.

EHRs have potential to eliminate problems associated with physical records such as loss and lack of accessibility. Can be stored centrally, accessible anytime, anywhere, irrespective of where and when the information was collected. Doctors can view patient's complete health record through EHR even if it is first time visiting. Reduces duplication of tests, improves quality, service efficiency and facilitates secure exchange of information which benefits patient and helps healthcare facilities to manage costs [10].

With EHRs generating massive amount of information, this can be used in different ways. Through Internet of Things (IoT) many devices such as health monitoring devices can produce huge volumes of information which requires big data processing so that it can be used by various companies.

Basically, big data in healthcare includes:

- Healthcare payer-provider data (EMRs, pharmacy prescription, and insurance records)
- Genomics driven (genotyping, gene expression data)
- Data from IoT.

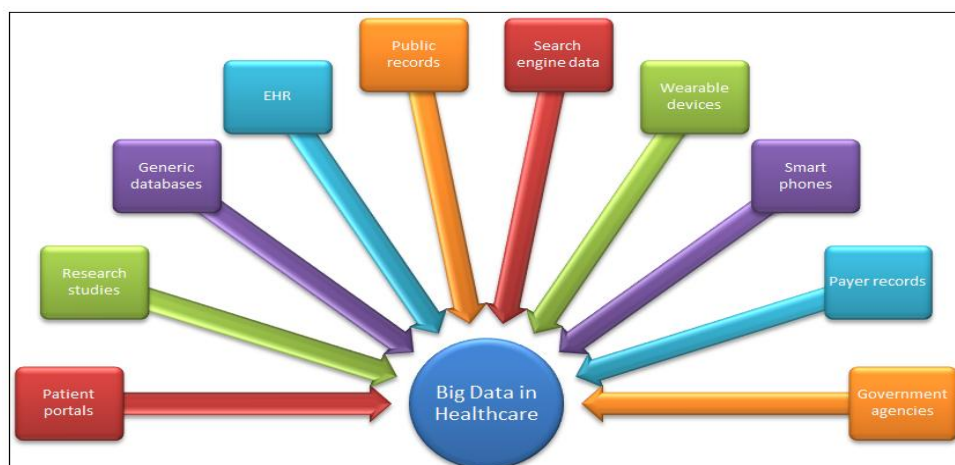


Fig-5: Sources of Big Data in Healthcare

Big Data Analytics

Data warehouses receive and store massive amounts of data generated from various sources. This is

then processed through analytic pipeline to obtain smarter and affordable healthcare options [2].

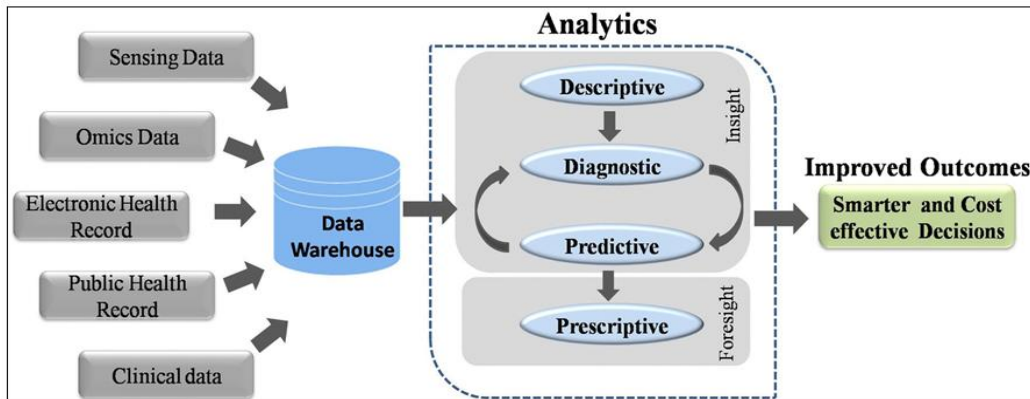


Fig-6: Big Data Analytics Workflow

Information analytics is the method of extricating, changing, stacking, displaying, and drawing conclusions from information to create choices. There are four ways of making sense out of information once it has been organized for detailing, and these are:

- Descriptive analytics – is designed to get you basic expository information: who, what, when, where, how many
- Diagnostic analytics – helps in answering the question of why something happened. Queries and drilldowns enable you to request more details which can help in explaining surprising values
- Predictive analytics – helps in identifying trends in relationships between variables, determine the strength of their correlation, and hypothesize casualty. Statistical modelling and predictive modelling go hand in hand
- Prescriptive analytics – AI and big data come into play. ML is about predicting outcomes based on numerous variables

Expressive and symptomatic analytics assist you build a story of the past whereas prescient and prescriptive analytics offer assistance in envisioning conceivable future.

Big Data Management

Huge information in healthcare faces challenges such as assortment of information from different sources and volume. The information is required to be put away in a record organize that's effortlessly open and lucid for proficient examination.

Another major challenge in healthcare is execution of high-end computing instruments, conventions, and high-end equipment in clinical setting. Different experts of science, IT, insights, and science are required to work together.

The information collected utilizing sensors can be made accessible on a capacity cloud utilizing instruments created. These apparatuses incorporate AI and ML advances to change over put away information into information. The most assignment is to explain, coordinated and display this complex information in fitting way for way better understanding. At long last, visualization devices created by realistic creators can be utilized to show this recently picked up information.

Heterogeneity of information is another challenge in healthcare. The foremost common stages utilized incorporate tall control computing clusters gotten to by means of lattice computing frameworks. Cloud computing is one such framework having virtualized capacity advances with solid administrations. It is solid, adaptable, and independent with omnipresent get to. Such stages act as recipient of information from omnipresent sensors to analyze and interpret the information in conjunction with reasonable web-based visualization.

The foremost common among different stages utilized for working with huge information incorporate Hadoop and Apache Start.

Hadoop



- The best consistent approach to analyze gigantic volumes of complex information is to disseminate and handle it in parallel on numerous nodes.
- However, the estimate is so expansive that thousands of computing machines are required to convey and wrap up preparing in a sensible sum of time.
- When working with thousands of hubs, one must oversee issues such as how to parallelize the computation, disperse the information, and handle failures.
- One of the foremost open-source conveyed applications for this reason is Hadoop.
- It executes MapReduce calculation for handling. It outline each coherent record in input into a set of middle of the road key/esteem sets, and decrease operation combines all the values that shared same key.
- It proficiently parallelizes the computation, handles disappointments and plans inter-machine communication over expansive scale clusters of machines.
- Hadoop Dispersion Record Framework (HDFS) gives versatile, proficient, and reproduction based capacity at different hubs that frame a portion of a cluster.
- It has other apparatuses that upgrade capacity and preparing components and consequently numerous companies like Yahoo, Facebook have received it.
- Many expansive ventures, like assurance of relationship between the discuss quality information and asthma confirmations, sedate advancement utilizing genomic and proteomic information, and other such viewpoints are actualizing Hadoop.
- Alternative to Hadoop, it may be a bound together motor for conveyed information preparing that incorporates higher-level libraries for supporting SQL questions (Start SQL), spilling information (Start Spilling), machine learning (MLlib) and chart handling (GraphX).
- These libraries help in expanding developers' efficiency as programming interface requires lesser coding efforts.
- Through Resilient Distributed Datasets (RDDs) implementation, in-memory processing of data is supported making Spark hundred times faster than Hadoop.
- Apache Start requires expansive sum of memory so fetched in higher than the difficult drive, MapReduce is anticipated to be more taken a toll effective.
- Apache Storm was created giving real-time system for information spilling processing.

Apache Spark



Table-2: Hadoop v/s Apache Spark

FEATURE	HADOOP	SPARK APACHE
Speed	Faster than traditional system	100x times than MapReduce
Written In	Java	Java, Python, Scala supported
Data Processing	Batch processing	Batch/ real-time/ iterative/ interactive/ graph
Ease of Use	Complex and lengthy	Compact and easier than Hadoop
Caching	Does not support caching of data	Caches the data in-memory and enhances the system performance
Framework	MapReduce	Generalized computation
Network	Distributed storage and computation	Distributed computation only
Storage	Usually data on disk (HDFS)	On disk/ in memory
Companies involved	Amazon Web Services IBM Infosys TCS	Yahoo Alibaba eBay Oracle Cisco

USING BIG DATA IN HEALTHCARE

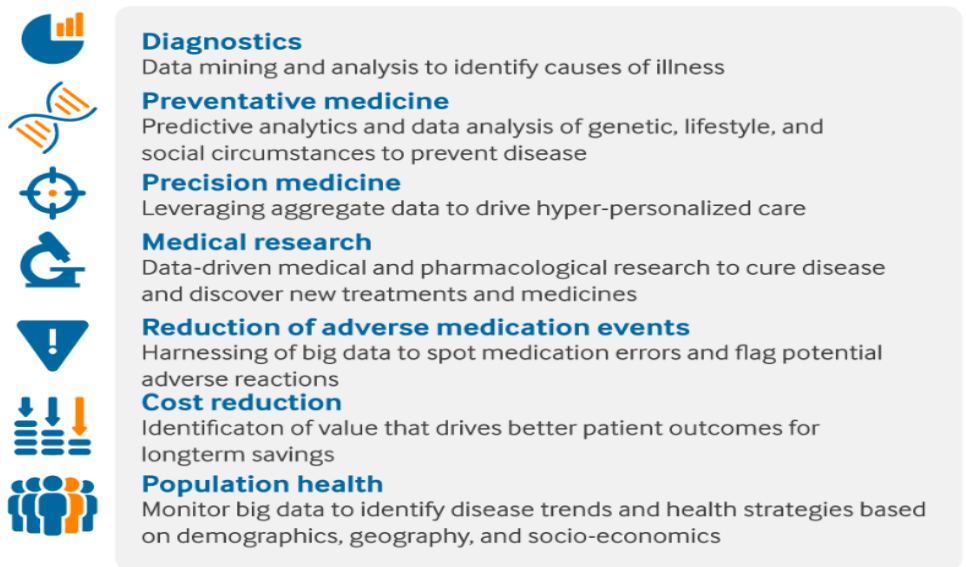


Fig-7: Applications of Big Data in Healthcare [11]

Clinical Operations

- To decide clinically pertinent cost-effective ways to analyze and treat patients through comparative inquire about
- Data collected from patients over time can help in enhancing quality of care and key choice making. This is improved through mechanized preparing of data through ML and faster to get huge volumes of information of a heterogenous structure [12].
- Israeli Start up MedAware is joining forces with healthcare organizations to send their choice back apparatus that employments huge information to spot medicine blunders some time recently they occur.
- The Mayo Clinic employments enormous information analytics to recognize patients with more than one persistent condition (comorbidity) as likely to advantage from early

mediations at care homes, in this manner sparing them from visits to the crisis office [11].

In Research:

- Big-data-driven drug development sets and strategies have been advancing with ever-expanding information from large-scale organic tests, clinical trials, and restorative records from members in information collection activities
- Predictive modelling to produce more targeted R&D pipeline in drugs and devices
- Improving clinical trial design and patient recruitment through statistical tools and algorithm to better match individuals to treatment, this reduces trial failures and brings new treatments to market soon

- Analysis of records of patients and clinical trials to identify indications that needs follow up and to discover any adverse effects before the product reach the market
- Improvement of pharmacovigilance and patient safety [13].

Public Health:

- Analyzing illness designs, following malady episodes and transmission moves forward open wellbeing observation and speed reaction
- Thorough development of more accurately targeted vaccines8
- Huge amounts of data can be converted to information used to identify needs, predict, and prevent crises for the benefit of the population

“Omics” studies:

- Next-generation sequencing (NGS) and Genome wide association studies (GWAS) generates data to decode human genetics [14].

- Scientist can study whole “genome” of an organism in “genomics” studies and similarly “transcription” of a single gene. This generates huge amounts of data with in-depth information than ever before
- NGS incredibly rearranged sequencing and the costs for producing entire genome grouping information. NGS innovation brought about in expanded volume of biomedical information from genomic and transcriptomic considers
- Combining genomic and transcriptomic data with proteomic and metabolomic data enhances our knowledge about individual profile of a patient – an approach often described as “personalized or precision health care” [2, 15].

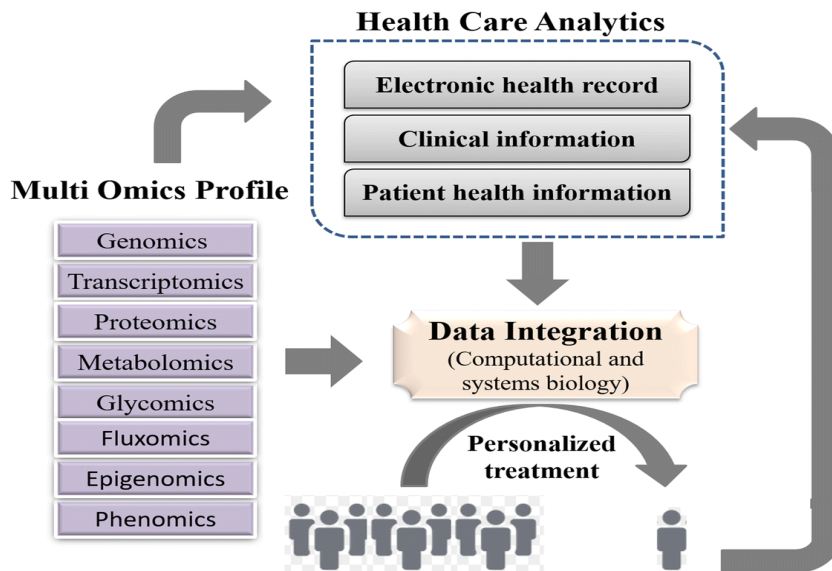


Fig-8: Personalized treatment

- There are numerous preferences expected from the preparing of “omics” information such as large-scale Human Genome Venture and other populace sequencing projects.
- In populace sequencing ventures such as 1000 genomes, analysts will have get to marvelous sum of crude data
- The Human Genome Project based Encyclopedia of DNA Elements (ENCODE) venture is pointed to decide all useful components in genome utilizing bioinformatics approach.
- Following are widely used bioinformatics-based tools for big data analysis on omics data,
 - SparkSeq – used for genomic data analysis with nucleotide precision
 - SAMQA – identifies errors and ensures the quality of large-scale genomic data
 - ART – can simulate profiles
 - DistMap – points to cover a more extensive run of sequencing applications. For case, one of its applications to be specific the BWA mapper can perform 500 million perused sets in approximately 6 hours, roughly 13 times speedier than a routine single-node mapper
 - SeqWare – enables access for large scale whole genome datasets by integrating genome browsers and tools
 - CloudBurst – used in genome mapping experiments to improve scalability of reading large sequencing data
 - Hydra – used for processing large peptide databases for proteomics datasets
 - BlueSNP – used for GWAS analysis
 - Myrna – provides information on the expression level differences of genes

IoT (Internet of Things):

- IoT makes a persistent stream of information whereas observing wellbeing of the individuals or patients, too known as Web of Restorative things (IoMT) or healthcare IoT, alluding to a worldwide foundation composed of restorative gadgets and applications interconnected through the Internet.
- These enable continuous observation and monitoring of health-related parameters such as
 - Sensors for physiological parameters (i.e., ECG, blood pressure, blood oxygen saturation, etc.)
 - Sensors for behaviour (i.e., pedometers, gyroscopes, GPS, etc.)
 - Patient context (i.e., relative air humidity, temperature, etc) [16].
- Technologies such as Radio Frequency IDentification (RFID) tags and Near Field Communication (NFC) devices enables objects to communicate and function as a web of smart things
- Capturing and analyzing real-time, voluminous, quick moving information from in-hospital and in-home gadgets for security checking and unfavorable occasion expectation (Raghupathi)
- Apps for portable gadgets, such as Aetna's Triage, prompt patients on their therapeutic condition utilizing amassed information and can suggest

patients look for therapeutic care based on input to the app.

- Apple has joined up with analysts at Stanford to decide in the event that the Apple Watch's heart sensor can be utilized to distinguish atrial fibrillation, a condition that causes the passing of around 130,000 Americans each year.

mHealth:

- mHealth is broader than fair versatile and smartphones, it incorporates wearables gadgets, inserts, location-based trackers, and sensors for bequest gadgets.
- Data from mhealth gadgets and big-data investigation can enable healthcare educate, giving a more all encompassing, patient-centric approach by concentrating on quality of treatment whereas adjusting weights on healthcare investing and cost.
- This is very beneficial to make quality healthcare affordable in rural areas.
- For illustration, MAiHEALTH gives portable and virtual therapeutic middle in rustic Patea. In Auckland, mHealth gadgets are supporting child wellbeing in lower socio-economic districts with the iMoko application. (Samaneh Madanian).

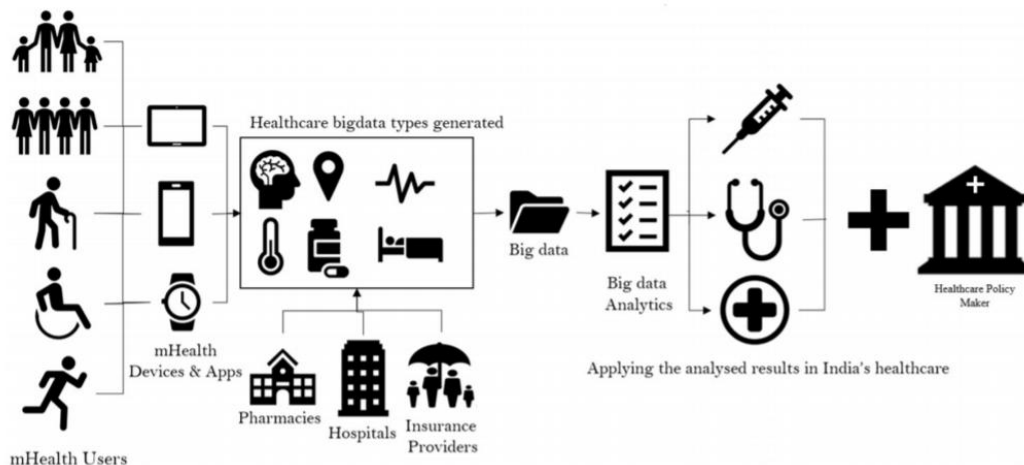


Fig-9: mHealth and Big Data

MACHINE LEARNING AND BIG DATA

Machine Learning (ML) dates back to the mid-twentieth century and can be characterized as an region of AL that employments computational calculations and insights to allow computers the capacity to “learn”, to make strides their comes about after preparing a adequate volume of data in a particular errand without unequivocal enlightening [12].

Records of patients contain recorded signals such as electrocardiogram (ECG), pictures and recordings together with content. This kind of organized and unstructured datasets have undiscovered riches of information which can be tackled utilizing progressed AI programs to draw bits of knowledge in quiet care.

ML or AI advancement makes a difference in compelling utilize of huge information in healthcare. It makes a difference to supply experiences through designs, affiliations, relationships and causations in complex, unstructured and unscaled datasets of huge information. Giving significant examination on datasets as shifted groupings of pictures utilizing Natural Language Processing (NLP) and bringing all these datasets together to produce expectation models, such as reaction of a quiet to a treatment regimen

Most of the information is in free text form (natural language), entered by professionals, AI techniques in healthcare can be used for automatic

creation of structured information from free text along with classifying and phenotyping of patients into different groups based on the contents of text written by professionals, these techniques are known as Natural Language Processing (NLP) [12].

Companies providing Big Data Analysis platforms in Healthcare Industry

To better understand the challenges of big data and for it to perform smoother analysis, many

companies integrated AI for analyzing results, text data and image data to generate meaningful outcomes.

In recent times, several companies and start-ups also emerged to provide healthcare-based analytics and solutions.

Table-3: Companies providing healthcare data analytics services

S. No	COMPANY	DESCRIPTION	WEBLINK
1	IBM Watson Health	Provides services on sharing clinical and health related data among hospital, researchers, and provider for advance research	https://www.ibm.com/watson/health/index-1.html
2	Medeanalytics	Provides performance management solutions, health systems and plans, and health analytics along with long track record facility of patient data	https://medeanalytics.com/
3	Health Fidelity	Provides management solutions for risks assessment in workflows of healthcare organization and methods for optimization and adjustment	https://healthfidelity.com/
4	Roam Analytics	Provides platforms for digging into big unstructured healthcare data for getting meaningful information	https://roamanalytics.com/
5	Faltiron Health	Provides applications for organizing and improving oncology data for better cancer treatment	https://flatiron.com/
6	Enlitic	Provides deep learning using large-scale datasets from clinical tests for healthcare diagnosis	https://www.enlitic.com/
7	Digital Reasoning Systems	Provides cognitive computing services and data analytic solutions for processing and organizing unstructured data into meaningful data	https://digitalreasoning.com/
8	Ayasdi	Provides AI accommodated platform for clinical variations, population health, risk management and other healthcare analytics	https://www.ayasdi.com/
9	Linguamatics	Provides text mining platform for digging important information from unstructured healthcare data	https://www.linguamatics.com/
10	Apixio	Provides cognitive computing platform for analysing clinical data and pdf health records to generate deep information	https://www.apixio.com/
11	Lumiata	Provides services for analytics and risk management for efficient outcomes in healthcare	https://www.lumiata.com/
12	OptumHealth	Provides healthcare analytics, improve modern health system's infrastructure and comprehensive and innovative solutions for the healthcare industry	https://www.optum.com/
13	Blue Health Intelligence	Its specialty is in building predictive models that support care management and identify at-risk patients to determine appropriate levels of care and disease management. BHI can even identify patient risk factors for rare conditions that enable clients to proactively address those factors.	https://bluehealthintelligence.com/
14	Information Builders	MedTech is recognizing the company for its Omni-Health Data platform, a suite of analytic and information management capabilities that offer payers and providers a comprehensive view of the healthcare practice.	https://www.informationbuilders.com/
15	Cardinal Analytx	Offers an AI-centric data platform that identifies people at high risk of rising cost and worsening health, then suggests interventions to prevent decline. Cardinal can also predict specific types of high-cost, high-acuity events well in advance, which provides the lead time needed to refer members to higher value providers.	https://www.prealizehealth.com/

IBM Watson

One of the greatest and experienced players in this division giving information investigation

administration is IBM Enterprise. IBM's Watson Wellbeing is an AI stage utilized to share and analyze wellbeing information among healing centers,

healthcare suppliers and researchers. It employs AL and ML based calculations to extricate most extreme data from negligible inputs. It coordinates a wide cluster of healthcare spaces to supply significant and organized data.

IBM Watson and Pfizer shaped a profitable collaboration to find novel immune-oncology combinations. Watson's profound learning modules coordinates with AI permits researchers to decipher complex genomic information sets. IBM with its

Watson Wellbeing computer framework has collaborated with Mayo Clinic, CVS Wellbeing, Commemoration Sloan Kettering Cancer Middle, and others. It has been utilized to anticipate particular sorts of cancer based on the quality expression profiles gotten from different huge datasets giving potential targets for drugs. It is additionally utilized in sedate disclosure programs by joining curated writing and forming arrange maps to supply a point by point outline of atomic scene in a particular illness demonstrate [2].

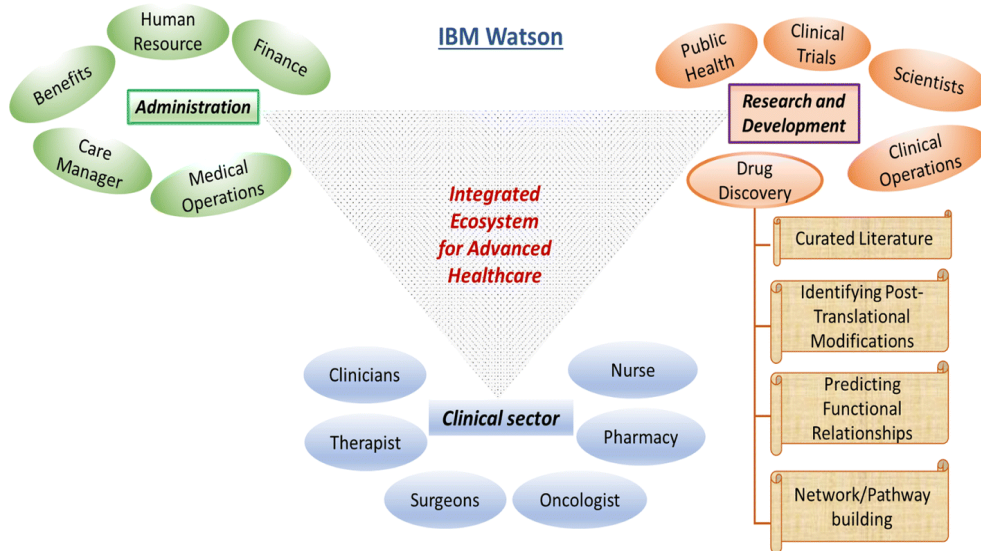


Fig-10: IBM Watson

Linguamatics

It is an NLP based calculation that depends on intuitively content mining calculation (I2E). Results gotten utilizing this procedure are ten times speedier than other apparatuses and does not require master information for information elucidation. This approach

can give data on hereditary relationship and actualities from unstructured information. NLP when coordinates in EHR or clinical records encourage extraction of clean and organized data that regularly remains covered up in unstructured input data.

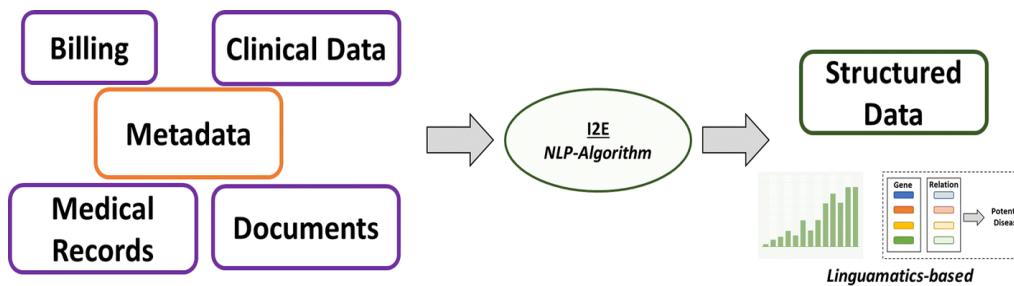


Fig-11: Linguamatics

AI and Big Data in COVID-19

AI along with Big Data has enormous potential, which is evident during COVID-19 and their role is anticipated to increase further in the future. It has been used from finding, tracking and diagnosing the

spread of virus in real time, facilitating public health interventions accordingly to repurposing old compounds and discovering new drugs along with potential vaccines [17].

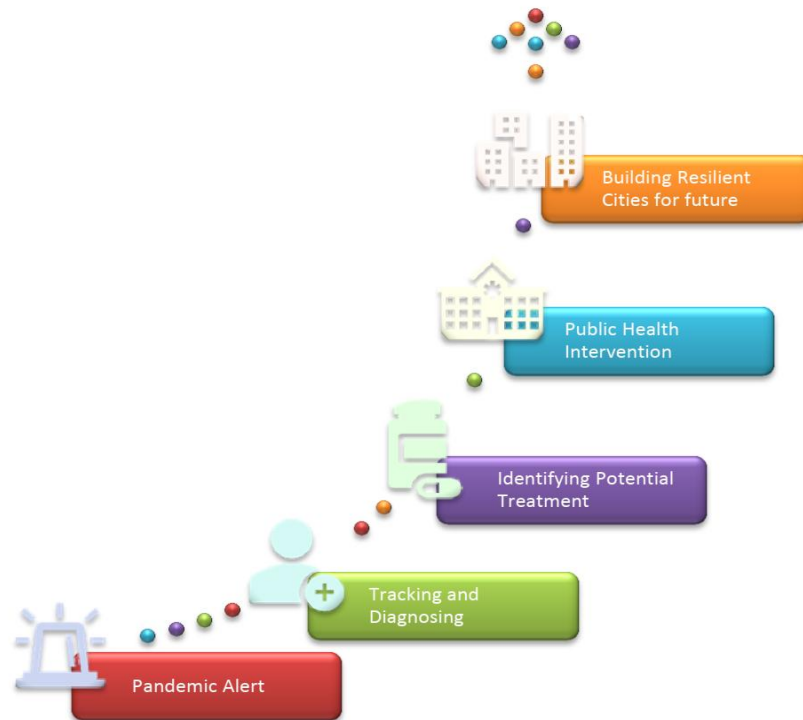


Fig-12: AI and Big Data in COVID-19

A fast and viable widespread alert:

- Big information can be utilized to foresee illness flare-up in real -time.
- With regard to past scourges and pandemics episodes, COVID-19 is uncommon in that open-access datasets containing day by day numbers of unused diseases broken down by nation, and, in a few cases, indeed cities, are broadly accessible.
- Combined with the data we have approximately the development of individuals; it speaks to the idealize dataset to combine scientific demonstrating and AI.
- Blue Dot, a Toronto-based start-up that employments an AI-enhanced reconnaissance framework, appears to have been the primary to identify the plague flare-up, a few hours after its insurgence within the to begin with detailed epicenter of Wuhan, well ahead of the Chinese specialists and other worldwide teach and agencies.
- Computational procedures empower us to imagine in real-time the spreading of the infection, such as the application outlined at the John Hopkins College, USA.
- Furthermore, enormous information collected from social systems and other related non-conventional information streams, empower us to reproduce early epidemiological story of the outbreak.

Tracking and diagnosing COVID cases:

- AI can encourage the conclusion of COVID-19 cases.
- For occasion, Infervision could be a start-up that utilizes profound learning restorative imaging

stages for encouraging fast conclusion of COVID-19 cases through the acknowledgment of particular lung features.

- Block-chain innovation could be a interesting decentralized framework of recording, confirming, and endorsing information and carrying out a arrangement of exchanges. It is characterized by a tall level of security and empowers the conveyance of patient-centered healthcare administrations, upgraded open wellbeing observation, administration of episodes and a fast and viable decision-making process.
- A low-cost block-chain and AI-coupled self-testing and following framework has been proposed for overseeing the COVID-19 pandemics.

Identifying a potential pharmacological treatment:

- Currently, no authoritatively endorsed helpful alternatives exist for the treatment of COVID-19. Doctors ought to give patients with steady management.
- Given the crisis of the circumstance creating a particular helpful would be time- and resource-consuming, researchers begun investigating three major approaches: (i) the possibility of utilizing as of now existing broad-spectrum anti-viral drugs, (ii) the plausibility of modifying/adapting them, and (iii) abusing accessible pharmaceuticals (either having a place to Western or Chinese conventional medication) for other helpful purposes (medicate repositioning or repurposing)

- AI can offer assistance rapidly distinguish potential therapeutics and candidate antibodies.
- For occasion, a collaboration between an AI start-up, BenevolentAI, and a college institution, the Royal College London, UK, has driven to the disclosure that baricitinib, a biologic utilized for the treatment of modestly to extremely dynamic rheumatoid joint pain in grown-ups, inhibitor of the janus kinases JAK1 and JAK2, may apply anti-viral effects.

Another start-up, Insilico Medicine, located in Hong Kong, has reported to have found six new drugs that may inhibit viral replication,

Facilitating the Implementation of Public Health Interventions:

- A start-up, Megvii, has reported the advancement of modern body and confront location and double detecting through infrared cameras and obvious light, as warm scanners for quickly screening subjects transiting in a swarmed put and distinguishing fever and tall temperature, possibly related to COVID-19.
- Ant Financial Services Group, once known as Alipay, of the Alibaba group, have formulated AI-based applications, based on parameters like self-reported wellbeing status, history of voyages and contacts, that can recognize COVID-19 cases.

- Environmental (contamination, squander transfer, climate alter) and wellbeing (communicable infections, such as nourishment- or vector-borne contaminations) challenges speak to genuine concerns that will debilitate citizens' liveability and quality of life.
- A profoundly coordinates frameworks approach is of vital significance, in that a sound city is like a living, energetic living being.
- Smart interconnected gadgets, wearable sensors and other imaginative innovations like savvy phones, broadcast communications, systems and GPS, empower to gather an extraordinary riches of information in genuine time that are user-centred and technology-driven, permitting: (1) quality observing, straightforwardness and responsibility, (2) asset allotment optimization, (3) citizen cooperation and incorporation, and 4) flexibility and adjustment to exogenous events.
- AI and Huge Information can give nearby choice- and policymakers with educated, evidence-based forecasts.
- Scientists have examined the COVID-19 episode embracing an urban point of view, appearing how keen cities and savvy systems can misuse high-quality, improved standardized conventions for information sharing amid crises, to superior oversee such circumstances [17].

Building Smart, Health, Resilient Cities:

CHALLENGES WITH BIG DATA



Fig-13: Challenges with Big Data

Data Structure:

- Majority of the data is unstructured in healthcare. It is often fragmented, standardized rarely, generated with incompatible formats
- EHRs don't share well inside organizational lines and

unstructured information is troublesome to total and analyze

- Poor quality of information, failure to coordinated information and need of believe within the information by suppliers and people contributing the information stay challenging [18].

- Data is more heterogenous in healthcare and are of unprecedented size
- Fragmentation emerges when EHRs are incapable to communicate viably between each other, locking understanding data into a restrictive framework [8].
- Shift to value-based installments has however to be adjusted, this limits speculation in endeavors to guarantee exact, strong clinical information are captured in EHRs

Data Standardization

- Data put away in different groups that are contradictory with all applications and advances limits interoperability
- Globalization of data results in dealing with variety of standards and differing terminologies
- Complexity from numerous sources, numerous databases and particular informatics dialect to get to the source of data [19].
- Unified format stands as a challenge for big data [2].

Data Storage:

- Acquisition and storage of big data, specialized needs for this also exists [20].
- Costs associated with storing and securing remain high
- With decreased costs and increased reliability, cloud-based IT infra is better which most organization opt for

Technical Expertise:

- Complexity of this platform requires a technical expertise to use big data to its full potential [20].
- Health care workers ought to be up to date with the utilize of continually changing innovation, proc

edures and continually moving standard of care [21].

- Due to steady advancement in innovation, there's a ought to bring clinician instruction into the 21st century empowering them to be able to connected with information researchers in building models commonplace around the world [19].

Data Governance

- Lack of controls, frameworks to oversee possession and duties for enormous information may be a challenge [21].
- Data administration is missing in consistency and thoroughness, inside and over organizations [18].
- Organizations ought to be mindful of different lawful issues emerging in prepare of overseeing tall level delicate information

Privacy and Security:

- Security and Privacy in big data are imperative issues.
- Security and privacy are not same.
- Although security is crucial for ensuring information, but it is deficiently for tending to privacy [22].
- Privacy concerns remains a priority in health care data analytics with enactment of Health Insurance Portability and Accountability Act (HIPAA) [18].
- HIPAA rules guide organizations with storing, transmission, authentication protocols and controlled access, integrity, and auditing
- There are significant concerns over confidentiality
- As information is centralized, it is highly vulnerable to attacks [21].

Table-4: Difference between Security and Privacy

SECURITY	PRIVACY
Security is the “confidentiality, integrity and availability” of data	Privacy is appropriate use of the information
Various techniques like Encryption, Firewall etc. are used to prevent data compromise	Organization cannot sell or disclose patient’s information without their consent
May provide confidentiality or protect an enterprise or agency	Patient’s right to safeguard their information from any other parties
It offers the ability to be confident that decisions are respected	Ability to decide what information of an individual goes and where to

CONCLUSION

With the adoption of Electronic Health Records (EHRs) in various countries, there is huge amount of patient data being produced. Along with EHRs, -omics data and data from IoT will result in Big Data repository of individuals and patients. With such database there exists a lot of opportunities for pharma industries, hospitals, physicians, and other stakeholders of healthcare industry to provide efficient and effective care to the patients. Many technologically advanced companies and healthcare industries are collaborating to

enhance quality care for patients at low cost. With emerging treatment options such as personalized treatment, data plays a key role. COVID-19 stands as the evidence for its potential when combined with technologies like AI and ML to identify drug treatments immediately available for patient care. But challenges such as data privacy and governance must be addressed to completely utilize its potential. Hadoop and Apache Spark are widely used for data management with many other technologies being developed for better outcomes. Since various data sources are available, standardization

and structure of the data collected acts as a barrier. This pandemic is an indication for healthcare industry to invest in much more technological advances and combine healthcare with technology to provide effective, cost-effective treatments in less time and Big Data stands as a healthcare technology evolution.

REFERENCES

1. Wang CJ, Ng CY, Brook RH. Response to COVID-19 in Taiwan: big data analytics, new technology, and proactive testing. *Jama*. 2020 Apr 14;323(14):1341-2.
2. Dash S, Shakyawar SK, Sharma M, Kaushik S. Big data in healthcare: management, analysis and future prospects. *Journal of Big Data*. 2019 Dec 1;6(1):54.
3. Powell A. Big data, massive potential. *Science & Technology*, 13 October 2015. <<https://news.harvard.edu/gazette/story/2015/10/big-data-massive-potential/>>.
4. Ristevski B, Chen M. Big data analytics in medicine and healthcare. *Journal of integrative bioinformatics*. 2018 May 10;15(3).
5. Gartner WB. Is there an elephant in entrepreneurship? Blind assumptions in theory development. *Entrepreneurship Theory and practice*. 2001 Jul;25(4):27-39.
6. Andreu-Perez J, Poon CC, Merrifield RD, Wong ST, Yang GZ. Big data for health. *IEEE journal of biomedical and health informatics*. 2015 Jul 10;19(4):1193-208.
7. Tom S, Dean F, Tod G. Electronic Health Records (EHR). *American Journal of Health Sciences*, 2014; 3. <https://www.researchgate.net/publication/267226700_Electronic_Health_Records_EHR>.
8. Agrawal R, Prabakaran S. Big data in digital healthcare: lessons learnt and recommendations for general practice. *Heredity*. 2020 Mar 5:1-10.
9. Srivastava SK. Adoption of electronic health records: a roadmap for India. *Healthcare informatics research*. 2016 Oct 1;22(4):261-9.
10. e-Health in India - Legal, Regulatory and Tax Overview. 2017. <http://www.nishithdesai.com/fileadmin/user_upload/pdfs/Research%20Papers/e-Health-in-India.pdf>.
11. Healthcare Big Data and the Promise of Value-Based Care. *NEJM Catalyst Innovations in Care Delivery*, 2018; 1.
12. Reiz AN, de la Hoz MA, Garcia MS. Big data analysis and machine learning in intensive care units. *Medicina Intensiva (English Edition)*. 2019 Oct 1;43(7):416-26.
13. Pastorino R, De Vito C, Migliara G, Glocker K, Binenbaum I, Ricciardi W, Boccia S. Benefits and challenges of Big Data in healthcare: an overview of the European initiatives. *European journal of public health*. 2019 Oct 1;29(Supplement_3):23-7.
14. Qian T, Zhu S, Hoshida Y. Use of big data in drug development for precision medicine: an update. *Expert review of precision medicine and drug development*. 2019 May 4;4(3):189-200.
15. Roca J, Tenyi A, Cano I. Paradigm changes for diagnosis: using big data for prediction. *Clinical Chemistry and Laboratory Medicine (CCLM)*. 2019 Feb 25;57(3):317-27.
16. S Rubí JN, L Gondim PR. IoMT platform for pervasive healthcare data aggregation, processing, and sharing based on OneM2M and OpenEHR. *Sensors*. 2019 Jan;19(19):4283.
17. Bragazzi NL, Dai H, Damiani G, Behzadifar M, Martini M, Wu J. How Big Data and Artificial Intelligence Can Help Better Manage the COVID-19 Pandemic. *International Journal of Environmental Research and Public Health*. 2020 Jan;17(9):3176.
18. Thompson ME, Dulin MF. Leveraging Data Analytics to Advance Personal, Population, and System Health: Moving Beyond Merely Capturing Services Provided. *North Carolina medical journal*. 2019 Jul 1;80(4):214-8.
19. Pépin JL, Bailly S, Tamisier R. Big Data in sleep apnoea: Opportunities and challenges. *Respirology*. 2020 May;25(5):486-94.
20. McPadden J, Durant TJ, Bunch DR, Coppi A, Price N, Rodgerson K, Torre Jr CJ, Byron W, Hsiao AL, Krumholz HM, Schulz WL. Health care and precision medicine research: analysis of a scalable data science platform. *Journal of medical Internet research*. 2019;21(4):e13043.
21. Kruse CS, Goswamy R, Raval YJ, Marawi S. Challenges and opportunities of big data in health care: a systematic review. *JMIR medical informatics*. 2016;4(4):e38.
22. Abouelmehdi K, Beni-Hessane A, Khaloufi H. Big healthcare data: preserving security and privacy. *Journal of Big Data*. 2018 Dec 1;5(1):1. <<https://catalyst.nejm.org/doi/full/10.1056/CAT.18.0290>>.