**Introduction to Medical Big data Analytics**

K.Sreelatha

Department of Electronics and Computer Engineering,

Sreenidhi Institute Of Science and Technology,Hyderabad, India sreelathak@sreenidhi..edu.in

## ABSTRACT

Big data is one of the largest and most complex datasets can be used to address business problems, medical prediction and diagnosis etc. Many fields now are facing tremendous challenges and various opportunities are provided to us to manage big data in various applications. Various big data tools are available to process the data and provide solutions to such challenges throughout the world. Here is a glimpse of such challenges faced and the major milestones that are achieved in the sizing of these data volumes to avoid information explosion.

.

**Keywords**—Big Data, Big Data Analytics, Medical big data.

## INTRODUCTION

The quest for Big data began seventy years ago when the first attempts was made towards quantifying the growth rate of *volume of data* better known as *“Information Explosion”* was predicted. In this section we will be discussing about the history of big data, challenges faced, its needs in today’s world and the various opportunities it gives us to manage big data in various applications. The major hurdles that one came across in the history of sizing data volumes and observations pertaining to data or information explosion are also a part of this discussion..

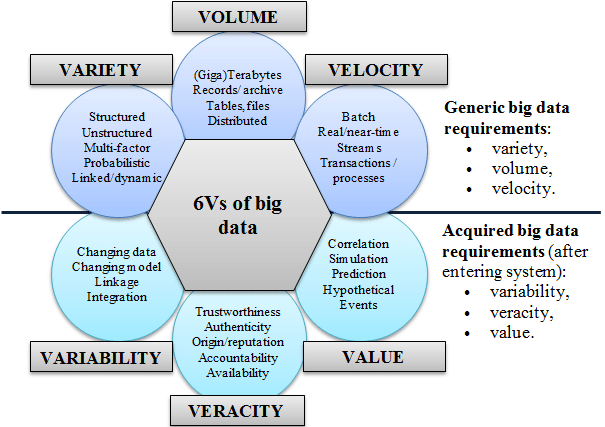
1. **Big Data over internet**
2. **Age of Big Data**

Super computers sites and large data centers must provide high–performance computing services to huge number of internet users concurrently. Data centers have to be upgraded using fast servers, storage systems and high-bandwidth networks. Data Lakes can be used to store data on cloud. The purpose is to facilitate storage for such large data stored into various data centers.

**C.Big Data – Definition**

Big data analytics can be seen from numerous angles. The business perspective, which includes value and commercial outcomes, the technological perspective, which includes managing computer resources and IT infrastructure, and the social perspective are arguably the three that are most frequently discussed [8]. The components of these viewpoints typically include both functional and non-functional criteria [9].

Figure 2 illustrates the various ways that big data influences both the strategy generation process and the strategiesthemselves, sometimes referred to as the "6V's"[10].



**Figure 2. 6Vs of big data**

**[Courtesy of Demchenko, De Laat and Membrey, 2014]**

**The V’s of the Health Care Big Data are:**

1) **Volume:** The healthcare data now includes wearable medical devices. These gadgets have the capacity to continually track a variety of physiological data, including bio potential, heart rate, blood pressure, and other parameters [2,4].

2) **Variety**: The complexity of various data types and the range of data sources could be used to describe healthcare data. In the past, paper prescriptions, handwritten notes, office medical records, MRI and CT scans, and other sources have typically provided the majority of unstructured data[5].

3) **Velocity**: Scripts, x-ray films, and paper files must be handled quickly[6].

4) **Veracity**: Healthcare data has biases, noise, and irregularities, which could jeopardize patients' access to appropriate decision-making and care [7]. "However, today, we rely on the general's trained eye."

5) **Variability:** It specifies how varied the data is? Data which is collected from different sources has data outliers to specify the difference of one type of data from another.

6) **Value:**  It specifies how efficiently the data is used and its utility in different applications .It also shows whether the data was useful for a specific application or not.

**D. Big Data –looking to internet as a model**

As we can see, data is expanding and coming from a variety of sources, sometimes unexpected ones:

Opening Data

• Social networks, where both public and private data are present.

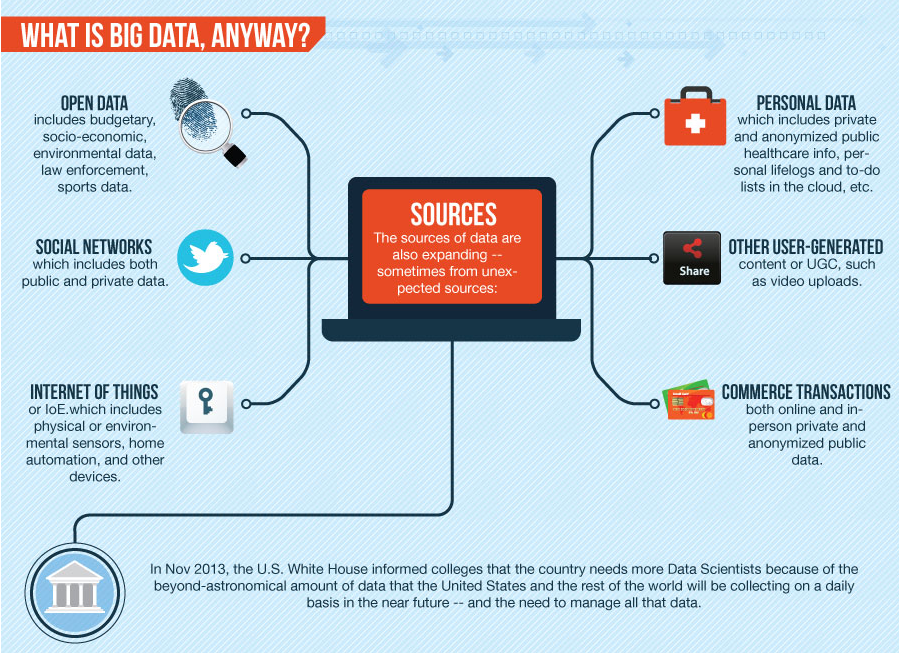
Internet of Things (IoT) (IoE)

• Individual data

• Additional user-generated content (UGC), like uploaded videos

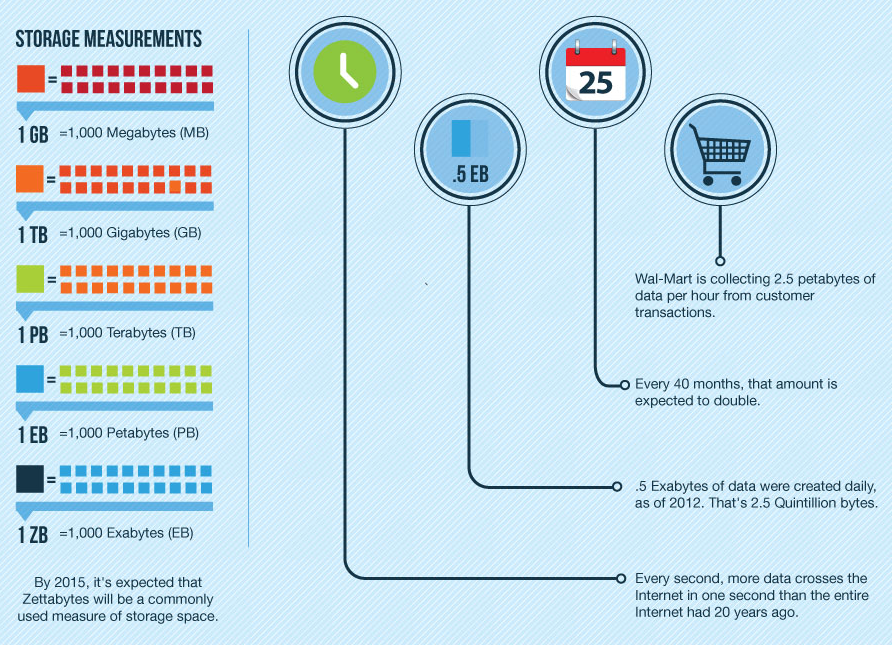
• Commercial dealings

The primary source of this enormous amount of data is the internet. Now the question is, "Are we prepared for the enormous volumes of data in the future? Figure 3 illustrates the sources of the data that were gathered and kept to help with this.



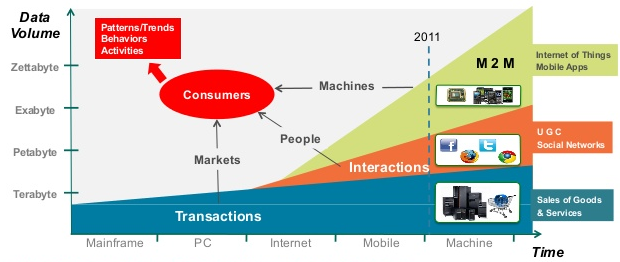
**Figure 3. Data collected from different sources and stored**

The rate at which we are collecting is expanding daily .Big data will come from anywhere and everywhere, and there is an immense need to access and consume data .The data is saved into any private cloud storage .Figure 4, shows how the data is collected and why it is collected? Now the big question lies on how will the data be managed?



**Figure 4. How much data is collected and why?**

There have been significant sources that include Internet of Things (IOT) and all the “smart” devices connected to it. In Figure 5, the graph of time v/s the data volume from different sources shows how the data is stored at different levels.



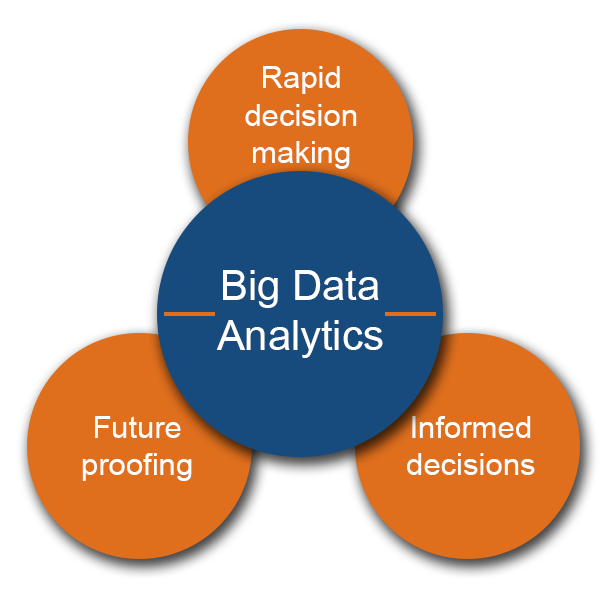
**Figure 5.Big data in consumer context**

**(Courtesy IDC and Berkeley Data growth estimate)**

**III.Big data analytics**

**A. Big data Analytics –Technologies for business processes**

  Big data analytics is the process of examining large and varied data sets to uncover hidden patterns, customer preferences and other useful information that can help organizations make more-informed business decisions as shown in the figure 6.



**Figure 6. Big data Analytics**

* Data scientists, statisticians, predictive modelers and other analytics experts can examine the increasing amounts of structured transaction data and other types of data thanks to big data analytics software. High-performance analytics systems are used in sophisticated applications involving elements like what-if studies, statistical algorithms, and predictive models.

**B. Big Data Analytics: Technologies and Tools**

* Unstructured and semi-structured data kinds often don't fit well in conventional data warehouses built on relational databases that are focused on structured data sets. In addition to Hadoop and its supporting tools, many enterprises that gather, process, and analyze big data use No-SQL databases. These include the following tools:
* YARN, a cluster management system that is a crucial component of second-generation Hadoop.
* MapReduce is a software architecture that enables programmers to create applications that analyze enormous volumes of unstructured data concurrently on standalone computers or distributed clusters of processors.
* Spark: An open-source framework for parallel processing that enables users to run complex data analytics programmes across clustered systems.
* HBase is a key/value data store that is column-oriented and built to work on top of the Hadoop Distributed File System (HDFS).
* Hive: a system for querying and analyzing massive datasets stored in Hadoop files.
* Kafka is a distributed publish-subscribe messaging platform created to take the role of conventional message brokers.
* Pig is an open-source application that provides a high-level framework for parallelizing Map Reduce tasks for Hadoop cluster execution.

**B. Uses and difficulties of big data analytics**

Microsoft and Amazon Web Services (AWS) have simplified things. for establishing and controlling Hadoop clusters in the cloud. Clusters are now created by users on the cloud, used for however long is necessary, and then shut down. Data silos can develop as a result of the usage of several platforms and data storage in a big data architecture, as well as from the volume and variety of data that is generally involved. These difficulties can affect data quality, consistency, and governance.

**C. Stakeholders**

Big Data is predicted to bring about a variety of motivations and goals for the various stakeholders in the healthcare sector, which can be summed up as follows:

1. Patients require patient-centered care.
2. Providers demand immediate access to patient information and other features.
3. Researchers want new tools to increase workflow quantity and quality.
4. Pharmacy need customer-friendly service.
5. Providers want real-time access to patient and other amenities.
6. Researchers want new tools to improve the quality and quantity of workflow .
7. Pharmacy companies want to better understand the causes of diseases.
8. Medical data centers consider safety monitoring and adverse event prediction [11].

**E. Opportunities**

  Healthcare organizations, from single-physician practices and multi-provider groups to huge hospital networks and accountable care organizations, stand to gain a lot by digitizing, merging, and leveraging big data effectively [12]. Clinical operations and R & D are two of the largest areas for potential savings with $165 billion and $108 billion in waste respectively [13]. Big data could help reduce waste and inefficiency in the following areas:

i) Clinical Operations

 ii) Research & development

iii) Public health

iv) Genomic analytics

v) Fraud detection

vi) Device/remote monitoring

**IV. ISSUES AND CHALLENGES**

When selecting a platform, factors including accessibility, usability, scalability, level of security, and continuity should be taken into account [16]. Data incompleteness, scalability, and security are some of the other difficulties with big data analytics [17,18]. Among the difficulties are:

* 1. Privacy and data security: The norms of the legal tradition of doctor-patient confidentiality govern the privacy of data specifically related to the healthcare industry. [34].
  2. Data is kept in formats that are incompatible with all technologies and applications [21, 24]. This lack of data standardisation also presents concerns with the transfer of that data [25, 26].
  3. Data Transfers and Storage: Data creation is less expensive than data storage. When data is produced, the expenses of safeguarding it rise and keeping them in storage remain high [26]. Both the act of transmitting data from one location to another and its analysis are costly [28,24,29]. Data scientists are still having trouble coming up with better techniques to process unstructured and heterogeneous data in the same way.
  4. Requirement of Appropriate Skills: The managing, storing, retrieving, and implementation of large data are difficult due to the knowledge in technical skills. Highly technical skill sets are required for data scientists. They must have soft talents like teamwork, leadership, creativity, and other things [27]

.**IV. Medical Big Data Analytics**

Information is the primary source of every system. Pervasive inefficiencies and a reward system poorly focuses on the key patient needs. They define a learning health care system as a system designed to generate and apply best evidence for the collaborative healthcare choices of each patient and provider.

1. **The Big Picture of Patient Data**

Any industry may now handle, analyze, and use data differently thanks to big data. Big data can be used to alter the healthcare industry. Healthcare analytics have the ability to lower treatment costs, foresee epidemic outbreaks, prevent infections, and more. Healthcare analytics ensure a superior result, i.e. healthy and secure patients.

**B. Why is big data used in healthcare?**

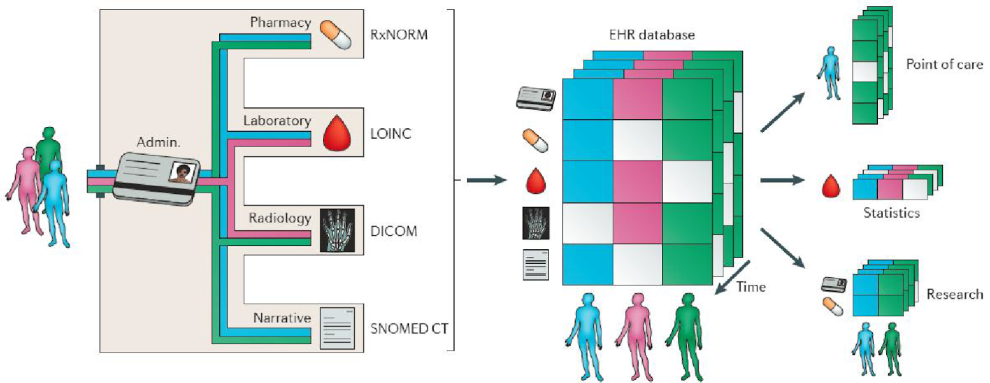
In recent years, it has become critical to use big data in healthcare and to control the rising expenses in countries like the United States. Medical judgments are becoming more and more evidence-based, which means they draw on a wide range of research and clinical data rather than just their education, and healthcare prices are significantly higher than they should be as well as expert opinion. Data management and gathering are becoming more important, just like in many other businesses.

**C.Data-Driven Healthcare Challenges**

There are numerous sources of medical data that are controlled by various states, institutions, and administrative divisions. Creating a new infrastructure would be necessary to integrate various data sources.

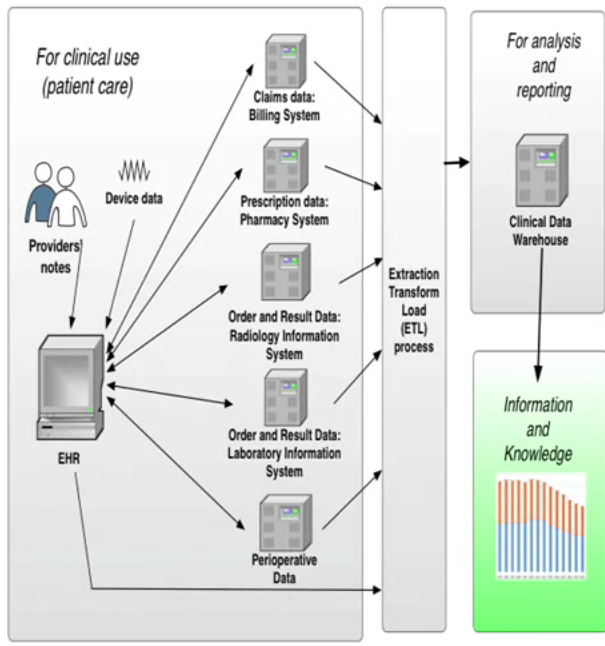
**D. Therapeutic Data Warehouse**

As seen in fig. 7, a clinical data warehouse compiles patient data into a single, well-organized place that is used for analysis and reporting.



**Figure 7. Electronic Health record of a patient.**

Since data can be stored in a variety of formats across systems, as shown in figure 8, ETL retrieves data from various clinical systems, synchronizes formats of data through a process known as transformation, cleans the data, and then imports the data into the database of the clinical data warehouse.

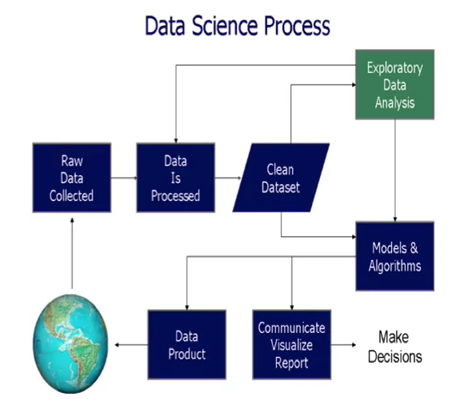


**Figure 8.clinical data warehouse**

Process of converting them to match is called transformation.

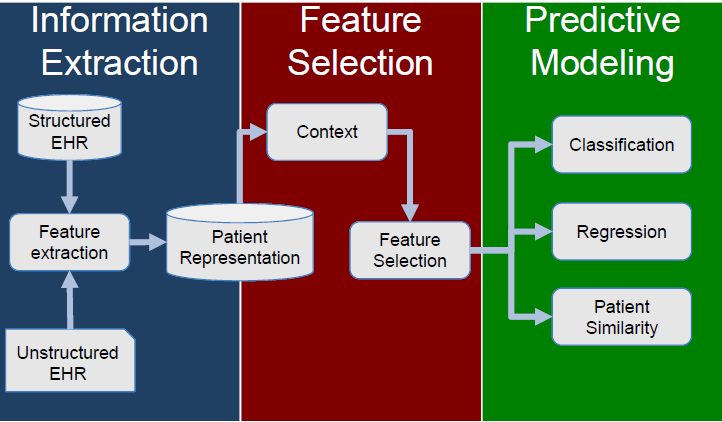
**V. Analytics**

1. **Process and Information Extraction**

The term analytics refers to the discovery of meaningful patterns in data, and is one of the steps in the data life cycle of collection of raw data, preparation of information, analysis of patterns to synthesize knowledge, and action to produce value. A centralized coordination is required to search the location for patient data for analysis and reporting. Analytical thinking is the synthesis of information into knowledge. One of the processes includes statistical analysis. as evident in the figure 9 below , one of the processes entails statistical analysis. As seen in Figure 9 below, raw data is gathered from a source, processed by exploratory data analysis, and sent for cleaning processed data to select the information needed. This information is then communicated with the aid of visualization tools like bar charts, pie charts, and other graphical representations.

**Figure 9. Process of Analytics**

Figure 10 shows the information extraction phase of the analytics process, where both structured and unstructured data are taken as input and features are extracted and stored as patient representations. Analytics is the entire process of data collection, extraction, transformation, analysis, interpretation, and reporting. The saved patient representation serves as the starting point for the feature selection step, which separates data according to the context in which it will be used. This output serves as the starting point for predictive modeling. Predictive modeling classifies data using three algorithms, i.e. Regression, classification, and patient similarity.



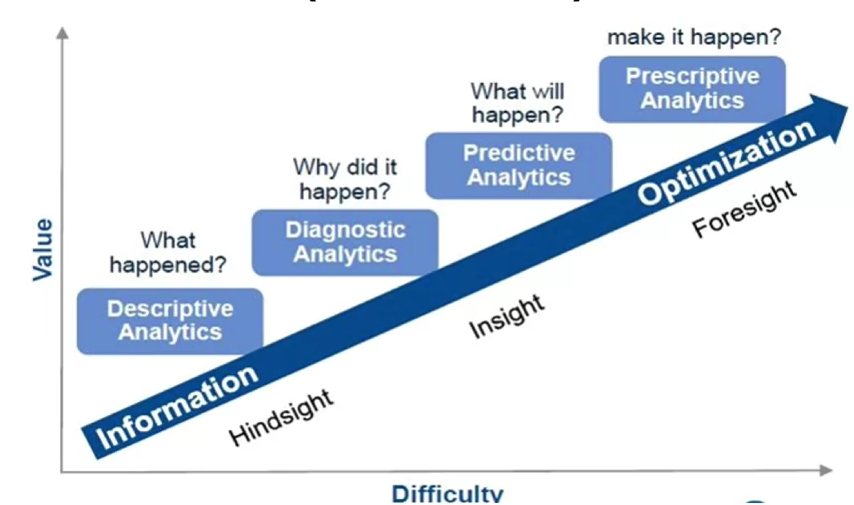
**Figure 10. Steps involved in Information Extraction**

**B.Types of Analytics**

In 2013, IBM divided analytics into three categories.

1.Descriptive: asks "what happened" using business intelligence and data mining.

2.Predictive: asks, "What might happen?" using statistical models and projections.

3.Prescriptive: asks "What should we do? " using simulation and optimization. ”Analyzes facts to determine "Why did it happen? The figure 11 depicts each stage of analytics.

**Figure 11. Types of Analytics**

Yet another query is, "Why did it occur? "is clearly explained. The diagnostic analytics are the most basic type of analytics, and they are the most important to the institute but also the easiest to execute. Predictive analytics are even harder but also more valuable. Prescriptive analytics are the final category and are both the most challenging and valuable.

The simplest sort of analytics, descriptive analytics merely explain the data. Common statistics include the number of laboratory tests, average patient age, and average length of hospital stay for patients with specific diagnoses. In descriptive analytics, visualization is frequently offered as tables, written narratives, bar or column charts, or pie charts.

**VI. Medical data analysis –Technical Issues**

The complexity of healthcare is influenced by the wide range of diseases and their co-morbidities, the heterogeneity of treatments and outcomes, and the different nuances of study designs, analytical tools, and methods for acquiring, processing, and interpreting healthcare data [6].

Because the extraction procedure makes raw data easier to understand, medical big data are usually collected following protocols (i.e., fixed forms) and are reasonably structured [9]. They might also be affected by a number of sources of uncertainty, such as measurement errors, missing data, or coding issues that alter the information contained in textual reports. As a result, domain knowledge may be crucial to both data analysis and result interpretation [10]. Additional aspects that set medical big data apart The different patient characteristic types, which may occasionally need to be weighted; the time structure, which may add another dimension; the information on the treatments received; and the time points at which decisions and changes to treatments are made (i.e., time-dependent confounding) [11] are analytical aspects.

A big data project's objective is to make sense of all accumulated data on as many distinct variables as is practical [12]. This is made possible by the rising availability and decreasing cost of computer technologies. Last but not least, big data technology is widely regarded as the key to a health system that is always developing and allows for two-way communication between research and operations [13]. Despite their many differences, clinical epidemiology and medical big data analysis share many parallels some of which are included in Table 1.

**Table 1. Medical big data analysis vs. classical statistical analysis**

|  |  |  |
| --- | --- | --- |
| **Parameters** | **Medical big data analysis** | **Classical statistical analysis** |
| Application | Hypothesis-generating | Hypothesis-testing |
| Questions of interest | Overcoming the limitation of locally or temporally stable association with continually updating the data and algorithm | Trying to prove causal relationships |
| Domain knowledge | More important in interpretation of the results | Important both in collection of data and interpretation of the results |
| Sources of data | Any kind of sources; frequently multiple sources | Carefully specified collection of data; usually single source |
| Data collection | Recording without the direct supervision of a human | Human-based measurement recording |
| Coverage of data to be analyzed | Substantial fraction of entire population | Small data samples from a specific population with some assumptions of their distribution |
| Data size | Frequently huge | Relatively small |
|  | Medical big data analysis | Classical statistical analysis |
| Nature of data | Unstructured and structured | Mainly structured |
| Data quality | Rarely clean | Quality controlled |
| Research questions of data analysis | May be different from those of data collection | Same as those of data collection |
| Underlying assumption of the model | Frequently absent | Based on various underlying probability distribution function |
| Analytic tools | Frequently automated with data mining algorithm | Manually by expert with classical statistics |
| Main outputs of analysis | Prediction, models, patterns identified | Statistical score contrasted against random chance |
| Privacy & ethics | Concerns about privacy and ethical issues | Data collection according to the pre-approved protocol; informed consent from the participants |

**A. How can medical big data be analyzed?**

The nine steps in the data mining process were listed by Iavindrasana et al. [19]: 1) learning the application domain, which includes determining the goal of the data mining application and pertinent prior knowledge of the domain; 2) dataset selection; 3) data cleaning and preprocessing; 4) data reduction and projection; 5) matching the goal defined in step 1 to a data mining method; 6) decision of the algorithm and search for data patterns; 7) pattern extraction; and 8) evaluation. The most often utilized performance metrics for evaluation include accuracy, sensitivity, specificity, receiver operating characteristic curve, precision, recall, f-measure, number of true predictions, and number of false positives [19].

**i) Missing value**

Three categories of missing data exist: entirely missing at random (MCAR), missing at random (MAR), and not missing at random (NMAR). MCAR is missingness whose likelihood is independent of the presence or absence of data. The likelihood of a missing observation is the same for all entities if the data are MCAR. The probability of MAR, or missingness, relies on the observed data. NMAR is missingness whose likelihood depends on unobserved information.

**ii) Curse of dimensionality**

The "curse of dimensionality" [6] refers to the challenges associated with optimization in high dimensional datasets [5] and has an impact on a variety of numerical studies, data sampling methods, combinatorial inference techniques, machine learning techniques, and data handling procedures.

Multicollinearity is unusual in large data analysis, while it may be frequent [6]. Model over fitting may be the cause of the generalizability problem. High dimensional data can be managed via feature selection [24] or dimension reduction [33]. Understanding that reducing dimensionality or feature selection may cause the loss of important mechanistic information is key. The benefits of learning new insights must be weighed against a false positive rate [25].

**iii) Bias control**

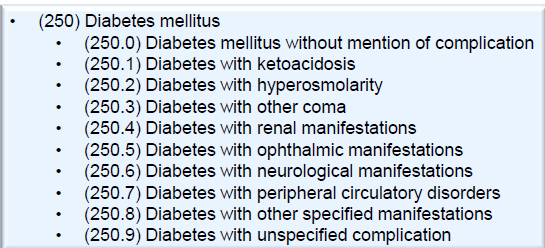
Randomized controlled trials are seen to be the gold standard of design validity because they minimize bias and control confounding [26]. A single randomized trial cannot be anticipated to deliver a gold-standard outcome that applies to all clinical investigations because trials frequently produce varied results [27]. Most significantly, research with insufficient power may draw invalid findings [7]. It is advised that unless these ideas are examined in sufficiently powered randomized controlled trials, the findings of observational research should not be used to guide clinical practice [28].

**VII.EHR (Electronic Health Records)**

It is a collection of patient and population health data that has been digitally saved. The sharing of these records is possible between various healthcare facilities. EHRs may contain a variety of data, including as billing information, demographics, medical history, prescription and allergy information, immunization status, laboratory test results, radiological pictures, and vital signs. EHR systems are made to accurately store data and to record a patient's state over time. It makes it unnecessary to locate a patient's previous paper medical records and helps to guarantee that the data is correct and readable. As there is just one changeable file, which increases the likelihood that the file is current, it can lessen the danger of data replication lost documentation EMRs are better at extracting medical data for the analysis of potential patterns and long-term changes in a patient since the digital information is searchable and contained in a single file. The extensive use of EHRs and EMRs may also make population-based research of medical records easier.

1. **Billing Data –ICD codes**

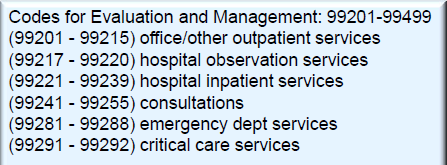
International Classification of Diseases is referred to as ICD. The World Health Organization maintains the ICD, a hierarchical terminology of diseases, signs, symptoms, and procedure codes (WHO). ICD-9 is most often used in the US, while ICD-10 is used worldwide..The figure below shows the clinical data of patients suffering Diabetes mellitus with their respective ICD codes which are universally accepted  as shown in figure 16.



**Figure 16. ICD codes of the patient suffering with Diabetes mellitus**

**b)CPT codes**

CPT stands for Current Procedural Terminology created by the American Medical Association. CPT is used for billing purposes for clinical services. The basic advantage is it shows high precision and the disadvantage is it has a low recall. The CPT code is as shown in the figure 17 below.



**Figure 17. CPT codes**

1. **Quality standards for medicines approved by FDA**

The Food and Drug Administration's (FDA) National Drug Code (NDC) is the industry standard; however, it is overly detailed, not adopted by all EHR systems, and differs between medications with the same constituents but different manufacturers. The National Library of Medicine's RxNorm is a normalized name scheme for both generic and branded medications. Typically, medications just store prescriptions, but we're not sure if patients completed those prescriptions or not.

Rich and varied sources of information can be found in clinical notes. It can be difficult to manage clinical notes because:

d) Short, imprecise phrases

d) Abbreviations

e) Misspellings

f) Semi-structured data: laboratory findings, vital signs

**A. Types of EHR:**

**i) Structured EHR**

The sharing of health information between clinicians and patients across various doctors, care giving organizations, and healthcare systems is standardized through information exchanges that link population health patterns.

**ii) Unstructured EHR**

With unstructured health data, it would be necessary to open each and every doctor's note from the previous few visits, make note of things like your blood sugar level, and later review the trends of all these notes to determine whether or not you had improved over time.

**B.Advantages and Disadvantages of Structured and Unstructured Data**

Documentation strategies relying exclusively on coded data entry have considerable drawbacks, including clinician time for note completion, difficulties determining medical relevance, and information loss, despite evidence of good impacts such as document completeness and simplicity of invoicing. In order to evaluate medical evidence, make management decisions, and effectively communicate medical information, narrative must be more precise and thorough. This should shorten a patient's hospital stay and eliminate pointless tests.

**IX. Big Data Challenges in Medical big data**

In spite of the methods available in Analytics the main question still remains .The various challenges in health care to be addressed are:

1. Inferring knowledge from complex heterogeneous patient sources.
2. Leveraging the patient/data correlations in longitudinal records.
3. Understanding unstructured clinical notes in the right context.
4. Efficiently handling large volumes of medical imaging data and extracting potentially useful information and biomarkers.
5. Analyzing genomic data is a computationally intensive task and combining with standard clinical data adds additional layers of complexity.
6. Capturing the patient’s behavioral data through several sensors; their various social interactions and communications.

**X. Enhanced Treatment Methods with Big Data**

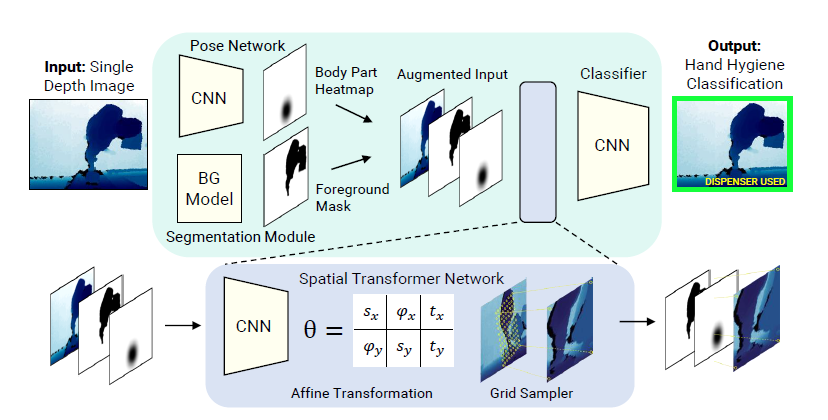
By concentrating and combining patient clinical records, one can create datasets for synthesis and then provide the data to others. The evolution of mobile applications is a current trend in digitization that supports patient healthcare. Anyone from anywhere can synchronize the patient's data using mobile apps for use in the latter. Big Data in healthcare plays a significant role in disease prediction, disease treatment, raising living standards, and preventing deaths. Treatment strategies are transforming swiftly in response to the world's population's rapid rise. Data is the driving force behind every change that these decisions entail. In the very near future, a doctor's consultation will even involve data exchange as a component of his or her diagnostic toolkit. even to access the vast, ever expandingdatabases, allowing issues with public healthcare to be identified and fixed before they materialize. Then, it is necessary to prepare well in advance for studying those disorders and identifying the best treatment options. The ground-breaking research, frequently including collaboration between experts in medicine and data, has the ability to look into the future and identify issues before they arise [17].

Because the data from public health sources is inadequate, there is a problematic curve when it comes to diagnosing the spread of infectious diseases. Real-time difficulties like disease data sources and imitable online and social networking platforms may be beneficial. Creating software to forecast information about common diseases effectively and researching social media's communication methods for effective forecasting of information on widespread diseases and studying how communication of social media is being used operatively.

**XI. New Generation of Digital Health Advisors**

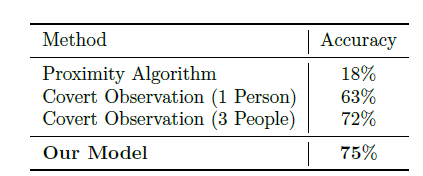
**1Computer Vision:** To improve operational efficiency and patient care, there has been considerable interest in building smart hospitals. The prevention of HAIs, or more specifically, the observation and tracking of hospital personnel hand hygiene, is one application for smart hospitals. RFID-based technologies are currently used to monitor hand hygiene. However, the resolution and precision of such devices are constrained. This example shows the efficacy of a computer vision-based method for tracking and monitoring hand hygiene compliance and identifies promising directions for future smart hospital research. Computer vision-based tracking systems have shown promising results in non-clinical applications such as self-driving cars and sports analytics.

A non-intrusive vision-based system is used in an experiment by the Stanford research lab [46] to monitor patient behavior in hospitals Both compliance with hand hygiene rules and the categorization of hand hygiene activities are measured. The interpretability of the system proposed is demonstrated through intuitive, qualitative results that examine human movement patterns and involve conducting spatial analytics. The hand washing activity is depicted in Figure 23 together with classifiers and transformation parameters.

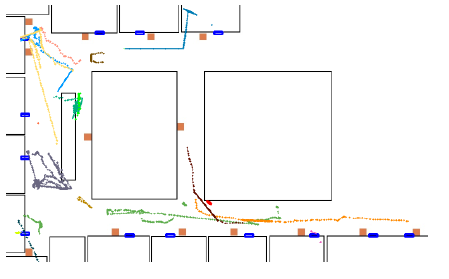


**Figure 23: Hand Hygiene activity classifier using various transformation parameters**

**3. Test Cases:** Data on actual hand hygiene compliance was gathered on a Friday between 12 and 1 pm. There were a lot of visitors in the hospital at (peak lunchtime). A total of 351 ground truth tracks were gathered and annotated; 170 of these tracks had someone entering a patient room, and 30 of those tracks were compliant (i.e., followed correct hand hygiene protocol). 34 tracks out of 181 tracks leaving a room were compliant. 150,400 photos make up the data set for the classifier, and 12,292 of those photographs showed people using the dispenser. As seen in figures 24 and 25, 80% of the photos were in the training set, while the remaining 20% were assigned to the test set.



**Figure 24.Comparison of hand hygiene**



**Figure 25.Top-down view of tracks**

Tracks seen from above: Black lines represent walls, orange squares represent dispensers, and entrances are indicated by blue rectangles. Different track colours represent various individuals. The approach described in this paper is an initial step toward vision-based smart hospitals and shows promising results for lowering Hospital acquired infections and ultimately enhancing patient care.

## XIII. Authors and Affiliations

K.Sreelatha, working as an Assistant Professor in the Department of Electronics and Computer Engineering, Affliated to Sreenidhi Institute Of Science and Technology, Ghatkesar, Hyderabad, India

Email id : sreelathak@sreenidhi..edu.in

## XIV.REFERENCES

[1]  I.A.T. Hashem, et al, “The rise of “big data” on cloud computing: Review and open research issues,” Information Systems, 2014.

[2] Jared Crapo. 2014. Big data in healthcare: Separating the hype from the reality. https://www. healthcatalyst.com/healthcare-big-data-realities. (2014).

[4] Bonnie  Feldman, Ellen M. Martin, and Tobi Skotnes. 2012. Big data in healthcare hype and hope. Technical

[5]<https://www.researchgate.net/profile/Tauseef_Naqishbandi/publication/271515514_Big_Data_CEP_and_IoT_Redefining_Holistic_Healthcare_Information_Systems_and_Analytics/links/54ca5ea70cf2c70ce521d024.pdf>

[6]  Report, Dr. Bonnie 360 (2012).

[7] 4 Kevin Hung, Yuan-Ting Zhang, and B. Tai. 2004. Wearable medical devices for tele-home healthcare

[8] Lake, P. and Drake, R. (2014). Information Systems Management in the Big Data Era. Cham: Springer International Publishing.

 [9] Pääkkönen, P. and Pakkala, D. (2015). ‘Reference Architecture and Classification of Technologies,  Products and Services for Big Data Systems‘. Big Data Research, 2 (4), pp. 166-186.

[10] Demchenko, Y., De Laat, C. and Membrey, P. (2014). ‘Defining architecture components of the Big Data Ecosystem‘. Proceedings of 2014 International Conference on Collaboration Technologies and Systems (CTS), Minneapolis, USA, pp. 104-112. Vol. 2. IEEE, 5384–5387.

[12] Bonnie Feldman, Ellen M. Martin, and Tobi Skotnes. 2012. Big data in healthcare hype and hope. Technical

[12] In 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEMBS’04) Report, Dr. Bonnie 360 (2012).

[14]Techcrunch. 2014. Healthcare’s Big Data Opportunity. http://techcrunch.com/2014/11/20/healthcaresbigdata-opportunity/. (2014)

[15] Bonnie Feldman, Ellen M. Martin, Tobi Skotnes (2012). Big Data in Healthcare Hype and Hope, online <https://www.ghdonline.org/uploads/big-data-inhealthcare_B_Kaplan_2012.pdf>.

[16] Burghard C: Big Data and Analytics Key to Accountable Care Success. IDC Health Insights; 2012

[17] Manyika J, Chui M, Brown B, Buhin J, Dobbs R, Roxburgh C, Byers AH: Big Data: The Next Frontier for Innovation, Competition, and Productivity. USA: McKinsey Global Institute; 2011.

[18] Manyika J, Chui M, Brown B, Buhin J, Dobbs R, Roxburgh C, Byers AH: Big Data: The Next Frontier for Innovation, Competition, and Productivity. 2011, USA: McKinsey Global Institute Google Scholar.

 [19] Data in Healthcare Hype and Hope, Bonnie Feldman Ellen M. Martin Tobi Skotnes,October 2012.

[20] Daniel A. Reed and Jack Dongarra 2015. Exascale computing and big data. Communications of the ACM 58, 7 (2015), 56–68.

 [21]  Wullianallur Raghupathi and Viju Raghupathi. 2014. Big data analytics in healthcare: Promise and potential. Health Information Science and Systems 2,1 (2014), 3.

 [22] Alexandros Labrinidis and H. V. Jagadish, “Challenges and opportunities with big data,” Proc. VLDB Endow. 5, pp. 2032-2033, August 2012 .

[24] Hsieh JC, Li AH, Yang CC. Mobile, cloud, and big data computing: contributions, challenges, and new directions in telecardiology. Int J Environ Res Public Health. 2013 Nov 13;10(11):6131–53.

[25] Chawla NV, Davis DA. Bringing big data to personalized healthcare: a patient-centered framework. J Gen Intern Med. 2013 Sep;28(Suppl3):S660–5. doi: 10.1007/s11606-013-24558.http://europepmc.org/abstract/MED/23797912.

[26] 25. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry.

[27]Overcoming Healthcare Big Data Challenges, retrieved from <http://www.mckesson.com/healthcareanalytics/healthcare-big-data-challenges/#footNote>.

 [28] Song TM, Song J, An JY, Hayman LL, Woo JM. Psychological and social factors affecting Internet searches on suicide in Korea: a big data analysis of Google search trends. Yonsei Med J. 2014 .

[29]Naqishbandi T, Imthyaz Sheriff C, Qazi S. Big data, CEP and IoT: redefining holistic healthcare information systems and analytics. Int J Eng Res and Technol. 2015;4(1):1 Jensen, Peter B., Lars J. Jensen, and Soren Brunak. "Mining electronic health records: towards better research applications and clinical care." *Nature Reviews Genetics* (2012).

[30] Senior M. Novartis signs up for Google smart lens. Nat Biotechnol. 2014;32(9):856. [[PubMed](https://www.ncbi.nlm.nih.gov/pubmed/25203024)]

[31] SleepBot [Internet] New York (NY): SleepBot; c2013. [cited at 2016 Jul 1] [32] RANKED Health [Internet] place unknown: publisher unknown; [cited at 2016 Jul 1].

[33] Auffray C, Balling R, Barroso I, Bencze L, Benson M, Bergeron J, et al. Making sense of big data in health research: towards an EU action plan. Genome Med. 2016;8(1):71.

 [34] Filkins BL, Kim JY, Roberts B, Armstrong W, Miller MA, Hultner ML, et al. Privacy and security in the era of digital health: what should translational researchers know and do about it? Am J Transl Res. 2016;8(3):1560–1580.

[35] Giraldo-Rodriguez L, Torres-Castro S, Martinez-Ramirez D, Gutierrez-Robledo LM, Perez-Cuevas R. Tele-care and tele-alarms for the elderly: preliminary experiences in Mexico. Rev Saude Publica. 2013;47(4):711–717.

 [36] MyTomorrows [Internet] Amsterdam: MyTomorrows; c2016. [cited at 2016 Jul 1]. Available from:[https://mytomorrows.com](https://mytomorrows.com/).

 [37] J. Tasioulas Big data, human rights and the ethics of scientific research ABC Religion and Ethics (November 30, 2017)

[38]C. Auffray, R. Balling, I. Barroso, L. Bencze, M. Benson, J. Bergeron, E. Bernal-Delgado, N. Blomberg, C. Bock, A. Conesa, *et al.*Making sense of big data in health research: towards an EU action plan Genome Med, 8 (2016), p. 71

[39]P.N. Schofield, J. Eppig, E. Huala, M. HrabedeAngelis, M. Harvey, D. Davidson, T.Weaver, S. Brown, D. Smedley, N. Rosenthal, *et al.*Sustaining the data and bioresource commons Science, 330 (2010), pp. 592-593

[40] J.L. Contreras Bermuda's legacy: policy, patents, and the design of the genome commons Minn J L Sci Technol, 12 (2011), pp. 61-125

[41]R.R. Faden, N.E. Kass, S.N. Goodman, P. Pronovost, S. Tunis, T.L. Beauchamp ethics framework for a learning health care system: a departure from traditional research ethics and clinical ethics European Commission General data protection regulation. Regulation (EU) (2016), p. 679

[42] SparSNP package http://bioinformatics.research.nicta.com.au/software/sparsnp/

[43] Wu, Tong Tong, et al. "Genome-wide association analysis by lasso penalized logistic regression."  *Bioinformatics* 25.6 (2009): 714-721.

[44] Akgul, Ceyhun Burak, et al. "Content-based image retrieval in radiology: current status and future directions." Journal of Digital Imaging 24.2 (2011): 208-222.

[45] Shang, Junliang, et al. "Performance analysis of novel methods for detecting epistasis." *BMC.*

[46] Terry Platchek, Arnold Milstein, Li Fei-Fei, Towards Vision-Based Smart Hospitals: A    System for Tracking and Monitoring Hand Hygiene Compliance, 1Proceedings of Machine Learning for Healthcare 2017 JMLR W&C Track Volume 68.