iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** XI **Month of publication:** November 2017

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue XI November 2017- Available at www.ijraset.com

Big Data: Healthcare Informatics

Purvika Shukla[1], Sonakshi Singh[2], Vandana S[3]
[1]*Student Computer Engineering Department Mukesh Patel School of Technology Management and Engineering*

Abstract: ***Collection, storage and analysis of health data have been, are and will be one of the cornerstone to provide efficient healthcare services and its importance is only multiplying considering the vast amount of health data collected every day. Knowledge discovery from various health data repositories requires the incorporation of healthcare data from diversified sources. While the healthcare industry is going digital ,the traditional and modern techniques of collection of data are facing mammoth challenges emerging due to the massive influx of data which is a direct result of population explosion. One thing which is certain in this ever-changing environment is that the advancement of technology goes hand in hand with monumental progress in the medical field. Revolutionary innovations and research have led to crucial strides being made in healthcare leading to life changing applications in providing diagnosis and treatments for diseases like cancer and chronic conditions such as hypertension to name a few. This paper covers the various hindrances faced by the healthcare industry in all in all its spheres and the possible solutions for the same.***

Keywords-Big Data, Data Mining, Health Data Warehouse, Healthcare Informatics,

I. INTRODUCTION TO BIG DATA

"There were 5 Exabyte's of information created between dawn of civilization through 2003,but that much information is created every two days -Eric Schema of Google in 2010". [21] The above statement in fact highlights the world we live in these days which is increasingly driven by data. How your organization reframes your data strategy and approach including big data and cloud technology will make a critical difference to performance in the future.

Big data defines holistic information management strategy that includes many of data and data management along with traditional data. The big data is defined by four Vs:

A. Volume

The amount of data. It could be from say a sensor from an equipment giving out data at tremendous speed, twitter data feeds etc.

B. elocity

This aspect relates to the speed at which data is received and acted upon.

C. Variety

The data could be structured or unstructured. The data could be video or audio which may require additional processing.

D. Value

Data has intrinsic value and you may have to apply quantitative and investigative techniques to derive the value.

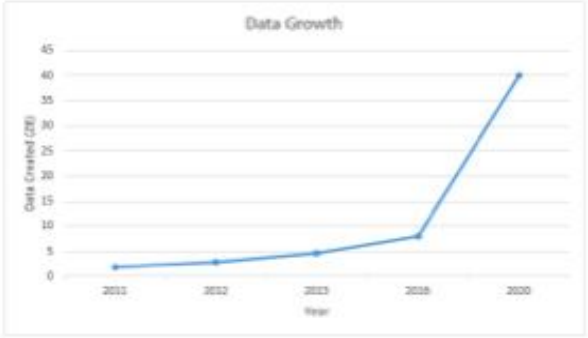


Fig 1.0: Growth of Data

There is a general perception that we are overwhelmed with data, making the ability to store, process, analyze interpret, consume and act upon data as a major challenge. For large scale, multinational organizations especially those in highly regulated industries such as finance, health care, the situation becomes more challenging. Traditional computing cannot match the data being generated and vital question is how to turn these data into actionable information. Digital and visualization capabilities are built into Big Data solution for achieving business growth by understanding complex data patterns. India has witnessed implementation of Big Data in the field of public health, FMCG (Fast Moving Consumer Goods), finance, banking etc.

II. A PEEP INTO THE HEALTHCARE INDUSTRY

The population of India continued to rise and there has been spurt in economic growth coupled with communicable diseases and 'lifestyle diseases' leading to the growth of health care industry. The industry in India in particular is expected to grow at
17 percent up to 2020 from 11 percent in 1990. As that happens, in rural areas, mobile technology and improved data services are expected to play a critical role in improving healthcare delivery. IT is set to play a big role with IT applications being used for social sector schemes on a large scale. In addition, several new computer and mobile phone based health initiatives have been launched such as e-Rakt Kosh a blood bank management information system.

III. ISSUES MAJOR CHALLENGES

A. Collection of unstructured data in healthcare and linking structured with unstructured data

Unstructured data is the information that typically requires a human touch to read, capture and interpret properly. It includes machine-written and handwritten information on unstructured paper forms, audio voice dictations, email messages and attachments, and typed transcriptions--to name a few.

Unstructured data in healthcare includes Patient- generated data from IoT devices such as fitness trackers, blood pressure monitors and weighing scales are also providing critical information about the day-to-day lifestyle characteristics of an individual. Insights derived from such data generated by the linking among EMR data, vital data, laboratory data, medication information, symptoms (to mention some of these) and their aggregation, even more with doctor notes, patient discharge letters, patient diaries, medical publications, namely linking structured with unstructured data, can be crucial to design coaching programmers that would help improve people's lifestyles and eventually reduce incidences of chronic disease, medication and hospitalization.

B. Maintaining record linkage during the integration of medical data is an important research issue

Healthcare data warehouses are highly beneficial in many fields such as mining health patterns, evidence-based medicine, personalized treatments, etc. Data required making proper medical decisions are trapped within fragmented and heterogeneous health systems that are not properly integrated. So the integration of these health records into a single warehouse is necessary. For maximum benefit from integrated health data repositories (IHDR), linkage of records is essential. Record linkage is the process of identifying record pairs from different information systems which belong to the same real world entity. [12] A solution to achieve this:

C. Patient identification technique based on secured record linkage (pitsrl)

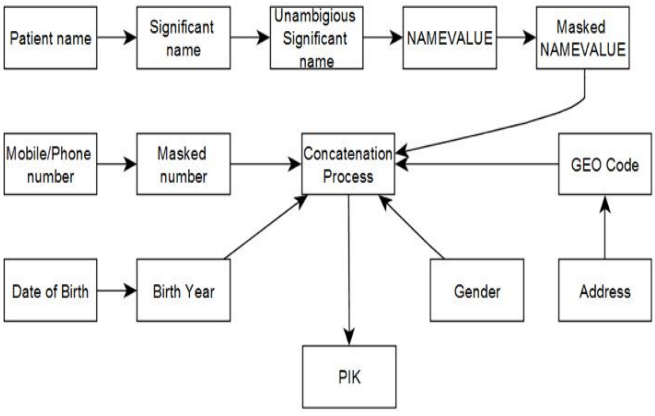


Fig.2. Block diagram of Patient Identification Technique based on Secured Record Linkage (PITSRL)

Health Records with Patient identifiable attributes such as name, address, date of birth in heterogeneous format from various health service providers are inputted in the systems. These records are then de-identified preserving record linkage. Privacy preserved linkable health records are then stored in national health data warehouse as a unified format. [12]The input of PITSRL system is health records provided by different healthcare organizations such as government and private hospitals, diagnostic centers, research centers, health NGOs. These data are in heterogeneous formats like Oracle, MS SQL or MySQL databases; CSV or MS Excel files, etc. In step 1, a Patient Identification Key based on Secured Record Linkage (PIKSRL) is generated for each patient record using available patient identifiable data. In step 2, all identifiable data, capable of identifying individual patients is removed from the health record We have used five attributes to generate identification key; mobile number, name, age, geocode, and gender. Mobile numbers are made secured through masking. Name is converted to NAMEVALUE. Age is used to generate the year of birth and age group. NAMEVALUE is the encrypted text string generated by our developed Name-Value Algorithm using significant and unambiguous characters contained in a patient's name.

D. Algorithm

1) Input: Health record set with patient identifiable data

2) Output: De-identified Linkable patient record

E. Steps

Repeat

1) Encrypt mobile number

2) Convert patient name to NAMEVALUE

3) Convert Date of Birth or Age to BIRTH-YEAR

4) Generate Patient Identification Key based on Secured Record Linkage (PIKSRL) from masked mobile number, NAMEVALUE, Birth year and Gender

5) Add PIKSRL Key to record set

6) Delete Patient name, Mobile number, Address, Date of Birth, Credit Card number data.

7) Until last record

F. Good Quality Healthcare

It is increasingly challenging to provide good quality care at a reasonable cost to their citizens when it is needed. The concept of the Iron Triangle of Health Care is often quoted to describe this very challenge. The three components of the triangle are quality, access and cost. Efficacy, value and outcome of the care reflect the quality of a healthcare system. Access describes who can receive care when they need it. Cost represents the price tag of the care and the affordability of the patients and payers. The problem is that all the components are typically in competition with one another in the Healthcare sector. [11] Thus while it may be possible to improve any one or two components, in most of the cases this comes at the expense of the third is shown in Figure 3

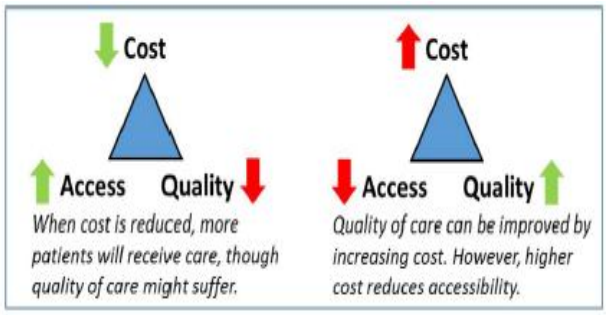


Fig.3. The examples indicate how current approaches Healthcare improvement often leads to sub-optimal solutions

G. To overcome the problems of missing data, ambiguity in patient identification, and high amount of noise in patient information.

These are the major challenges because the healthcare data is growing exponentially every single day and nearly all the data is being stored in separate silos. Therefore deriving insights and value from the aggregation of these data sets is not possible at this stage.The current need is to be able to maintain a privacy preserved record linkage during the integration of medical data.

H. Issues of Patient Generated Health Data (PGHD) in clinical work flow

The reception that PGHD has received can be summarized as mixed. The concerns are from different components of the data flow structure so to speak.

1) *Provider Concerns:* Provider concerns deals with hesitancy to accept the rapid surge of PGHD as providers fear that it may hinder their ability to deliver quality care. Also another worry is the possibility of unrealistic expectations from the patient regarding their healthcare .Most importantly, providers are also reluctant to fully embrace PGHD due to the very fact that they may be held accountable for information that might have not be reviewed in a timely fashion and the information that may require a urgent response[22].
2) *Patient Concerns:* Patients may be concerned about the failure of healthcare providers to meet the expectations of their healthcare. That is whether or not the doctors have reviewed their data , have tabulated it, confirmed the receipt, whether the information has been saved in the patient's chart , whether the information was securely received and stored and so on .
3) Technical Issues*:* The information must be collected and submitted in standardized ways that ensure that the information cannot only be received but also understood and integrated into the EHR if desired.
4) Privacy and Security*:* Authentication of patient needs to be ensured so that information can be assured to him or her with confidence. All stakeholders need to be assured that the channel of PGHD is safe and secure. Efficient linkage of information from the patient to Health provider is imperative for tracking data. Identifiable health data in electronic health repositories in electronic health repositories may produce a significant risk to patient privacy and also make the health information systems security vulnerable to hackers. Collecting and integrating healthcare data in developing countries for developing national health data warehouse. [12]

IV. PATIENT GENERATED HEALTH DATA

Patient Generated Healthcare Data, covers the foundations of the healthcare namely health history, the various symptoms, treatment history, biometric data, choices related to one's way of living and other such health related formation which is generated, assimilated and recorded in a way which can help address a health concern, has encountered a surge like never before with the influx of data that is available in today's tech savvy environment .Its inference can be through the means of the patients (both by and from patients)or their care partners who tend to their medical needs in a formal or informal set up.[11]Essentially, Patient Generated Data collects data in a unorthodox care setting beyond the traditional set up. In fact in the possible future, a majority of data necessary will be collected in these unorthodox settings. Health care organizations have started accepting some PGHD in recent years with its source being a far cry from the traditional clinical settings. However, progress is being made to attain a level of familiarity with the PGHD framework and structure. There is a flow to the collection-capture, transfer and review. In PGHD, data capture may refer to the creation and storage of health data by the patient or designee (provider) and entails written data inputted through a keyboard or other such input device, verbal data entered via a microphone and physiological and/or environmental data recorded on a monitoring device.

PGHD can help bridge gaps in information and can offer healthcare providers a way to track a patient's health status and compliance to a therapy in between medical visits. PGHD allows a way to gather information on a continuous basis rather than at a single point in time. Even though the healthcare industry is transitioning and becoming digitized which is a mammoth, challenges exist in integrating Patient Generated data to the clinical data workflow. Harmonization is the key. Even though it is revolutionizing healthcare as we know it, the concerns regarding it are also increasing

V. CLINICAL DATA AND ITS PROBLEMS

In Fact, the hurdles faced in the traditional clinical settings of the healthcare industry are also making their presence felt due to the volume of data, speed of the generation of this data and the complexity emerging out of multiple and non-standard formats. The main hurdle is definitely the occurrence of patient data residing in different systems often unlike and dissimilar. These separate systems contain duplicate information for the same patient. Consequently, this hinders the diagnostic process as the right information isn't available at the right time. One challenge that plagues the medical industry is the patient matching and identification. Without precise patient identification, portions of patient health record may be fragmented into isolated episodes (duplicates) or linked to the wrong patient. Documents that are stored and gathered in different clinical settings need to be matched and linked to the correct patient record. Patient information needs to be connected effectively so that the decisions made by clinicians are based on correct patient records. Minimizing duplicate information and overheads is imperative. Hence, the need to rectify healthcare framework and make it more productive has brought about the urge to identify, cleanse, match, de-duplicate

and merge patient records to create a master index that may be used to generate a complete and single view of a patient. By using various Big Data Analytic techniques like fuzzy matching algorithms, Map Reduce etc, an effort is being made to effectuate change in the healthcare industry as well as provide solutions to the problems

A. Lack of nationwide unique identifier

Without a nationwide unique patient identifier, to accurately match multiple patient records is a monumental challenge. However steps have been implemented in a bid to tackle this challenge, for eg the Enterprise Master Patient Index(EMPI) uses a variety of statistical algorithms to match patient records while striving to minimize the abundance of false positive and false negative matches [17]

Even with the existence of such systems there exists a huge amount of duplicate records.

B. Disparate Systems

Several Healthcare systems maintain their own data on individual platforms and there is rarely a sharing of information among several platforms. Due to nature of data captured in healthcare system, it is extremely difficult to hold information on a single view legacy platform. This lack of communication can have a negative impact of patient's care.

C. Scalability

Healthcare is turning digital and it is fragmented. The ever increasing volume of data has led to greater difficulty in handling of such large sets of data coming at high speed. Hence, the challenge of matching huge volume of patient records persists.

D. Costly and inflexible solutions

Existing toolsets for patient matching which are available in the commercial market are costly with high licensing fee. They aren't very flexible and are limited to a few fuzzy matching algorithms.

Proposed Solution of of Patient Matching using Big

E. Data Analytics

First, Patient Data is extracted from various different systems like clinical, Laboratory, pharmacy, social media etc. Data is integrated, cleansed and transformed and then patient matching is performed using fuzzy algorithms and Hadoop Map Reduce. After that data is segregated into matching patients and non matching patients

Fuzzy Matching technique involves matching of patients (Name ,ID, and Address) across disparate data sets via synonyms, phonetics and approximate spellings. Fuzzy matching is an advanced mathematical process in which the similarities between data sets, information, and facts are processed. The subsequent result of these similarities in not always true nor false or 100% accurate. Hence, the term fuzzy. In this, a data type of any length from any place in a field is compared to find non-exact matches. For each specific data entry examined, a probability score is generated using fuzzy matching process to determine the veracity and accuracy of the match. For example, the name Thomas A might get a 90% similarity score, while Tom A might get a 75% score. Similarity can be assessed by calculating distance between 2 pieces of data information, the more dissimilar are the two sets of data.

To tackle the enormity and the monstrosity of data and the speed at which it is incoming, MapReduce is implemented using Levenshtein Distance algorithm[17].Levenshtein distance is the distance between two words which is the minimum number of single character edits needed to make one word same as the other .MapReduce can be declared as a platform which is a programming model based on different distributed computations on huge amounts of data which includes execution framework for the processing function on huge clusters of commodity servers. So, the basic function of MapReduce is to match patients' data which basically entails multiple attributes of a patient identity namely Birth Address, City, Pin Code so on and so forth. Afterwards, each attribute identity is assigned an associated weight. These weights vary and are indicative of value(e.g Last name counts for more than a First name match, Date of Birth match counts more than a city match).For any pair of entities, distance is calculated between corresponding attributes and the distances are aggregated over various attributes of an entity to find the distance between two entities. With MapReduce, data about a patient identity is produced with various keys from Patient Name, Date of Birth, Gender, etc. These attributes are grouped and in reduce phase a distance with weight is calculated to find out whether any existing records are similar enough to be considered a duplicate of another. Several parameters or thresholds, to judge this distance are laid out. For instance, distance is 0 for identical records, distances close to 0 for records that are

identical with typographical errors. Records with are different will have values greater than 5.

VI. BIG DATA AS AN E-HEALTH SERVICE (BDEHS)

Digital Healthcare Solutions have the capability to transform the healthcare industry. Already almost 1000 petabytes of data have been generated through various channels such as Electronic Medical Records (EMRs), personal health records, genetic sequencing etc.[18]

EMR data is huge, but needs to be validated for meaningful analysis. Organizations supporting health infrastructure also provide external data. Data input to the BDeHs which is a spot light of this study includes structured, semi-structured and unstructured data, ranging from hospital data, insurance data and data found after research. Available big data technologies such as STORM, Map Reduce and Hadoop are used. The BDeHs provides services to access, organize and take out discoveries from huge volumes of e-health digital data. In order to be effective, key capitalism is identified in BDeHs system:

A. Data federation aggregation
B. Security and regulatory concerns
C. Data operational management

The lution to BDeHs, which provides a data streaming federation and decision points in supporting the original flow of e-health. A data flow in this e-health environment is mapped into a stream with additional processing stages. The solution also addresses security and regulatory compliance. Quality of Service generates real time processing, enhancement of processing cluster and network capacity, data interoperability management as well as reporting capabilities.

VII. LEVERAGING BIG DATA TO PROVIDE A WEB SERVICE THAT PROVIDES THE LIKELIHOOD OF DEVELOPING PSYCHOLOGICAL CONDITIONS AFTER A CONCUSSION

It has been highlighted that significant amount of data has been acquired from patients suffering from mild traumatic injuries and psychological health problems after a concussion. Advances in big data technology provides the opportunity to map the condition of the patient before injury and predict the likelihood of developing a psychological condition. A REST web service is used which uses the clinical data of patients who have suffered a concussion and predicts the likelihood of developing a psychological disorder. The concept of predictor psychological conditions using neural networks is used. The Web Service which uses neural networks sought to provide short and long term predictions post concussion. Hadoop is used along with Python for computing and running neural network, with integration support from Django.

The Web Service uses separate neural network for short and long term predictions. A user friendly front end user interface was also available. The user has an option to save their data to the database else the data is discarded post training against a pre trained network. Using the information provided by the web service, The clinician/patient can take preventive measures to combat these diseases.

VIII. HEALTH CARE FOR ELDERLY CITIZENS USING BIG DATA

“The paradox of life; everyone desires a fuller life. But no one wishes to increase in age.” It is said that India is a nation of youth and opportunities are plenty. But we have to keep in mind that as we embrace development, our population is set to grow. Fig 4 gives a trend of growth of elderly populations in India . The healthcare aspects of elderly citizens should be a focus area if we transit to the arena of developed countries . [13]

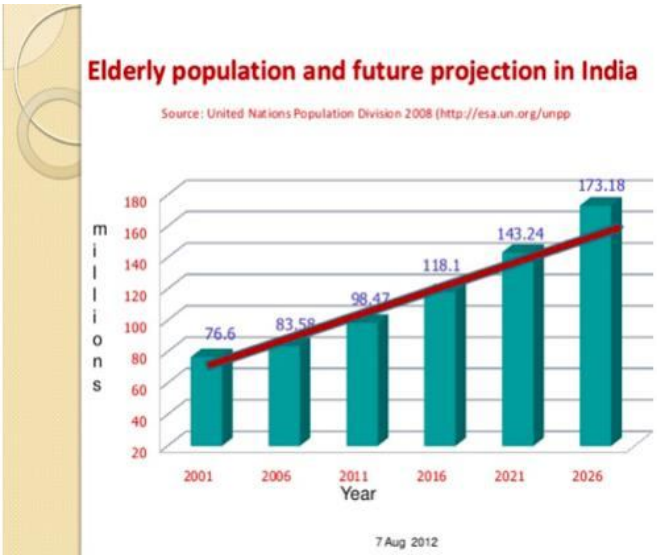


Fig .4: Population of Elderly Citizens – India

In India, while conservative rules still ensure elderly citizens stay with their family , however, as families get smaller more aged people live alone. Opportunistic business ventures have opened up retirement homes as new area of thriving business. There are exclusive residential areas for the senior citizens to move in, which provide wholesome support. Big data has a big potential in health care. This technology can be put to use to ensure that elders are monitored, diagnosed and medicated well in time to give them a healthy last phase of life.

A number of state of the art technologies can reduce cost and assist in better management of chronic diseases. These devices include ones that constantly monitor health indicators, instruments that auto administer therapies or ones which assist a patient in self administration of therapy. The internet and smartphones have pervaded human space to considerable extent and du to this mobile applications are now increasingly used along with Internet of Things (IOT).

The changeover to Electronic Health Records (EHR) has revolutionized the management of patient records. However, there is a challenge to overcome because the variety of sensors speak in different languages and are difficult to integrate, and coupled with this aspect there are concerns on privacy and people tending to hoard data .

The ICT will drive development of devices and apps to monitor health. The classic examples are :

A. The Myo , an original motion controller for games is being used for physiotherapy post orthopedic treatment.

B. The Zio Patch measures heart rate and ECG and is an approved by US FDA.

Let us say , Fitbit can provide huge amount of biometric data to researchers for testing hypothesis on nutrition, fitness, disease progression , success of treatment etc. But data scientists is of the opinion that the data requires cleaning and validation prior put to use. Say for example a problem called hypertension can a particular code. But is important to understand whether the hypertension was the original reported problem or was diagnosed during the investigation. So an algorithm developed on the big data after extensive iteration can present a solution which can prevent hospitalizations by identifying patients at greatest risk and enhancing their patient care before they are admitted into a hospital.

Wrist devices with sensors such as accelerometer for measuring activity of the wearer, temperature sensors to measure ambient and skin temperature and also sensors to measure heartbeat have been developed. The sensors have digital serial interface for data communication. The biggest opportunity in

respect to big data for elderly healthcare is type and source of data. The downside of growing size of big data is which presents challenge to the analytics is that 90% of data is not usable. However developments in cloud computing, visualization are helping development of platforms which can handle such tasks. [2]

One of main focus area for application of big data for healthcare would be their use for elderly citizens. Wearable devices and mobile applications are considered drivers in this field but would require strong algorithm based big data solution to avoid overwhelming of system due to irrelevant data and to generate meaningful clinical advice. The area has a huge scope for improving the availability of cost effective healthcare.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue XI November 2017- Available at www.ijraset.com

IX. BIG DATA'S INFLUENCE ON CLINICAL CARE FOR CANCER

The hypothesis is that clinical decision-making and outcomes in cancer care can be improved by the combination of datasets from real-time clinical information systems with historical datasets from clinical trials and the following application of big data and business intelligence techniques.[19] Clinical decision-making and the consequent result improvements is determined by clinician and patient feedback on usage, intervention changes and outcomes using structured interviews and surveys, augmented by changes in survival rates and patient welfare. However, there are problems in combining datasets as well such as different purposes of the concerned datasets, different patient populations, different primary keys, different data models, ensuring privacy, and proprietary issues.

While extraordinary potential exists to enhance health outcomes by consolidating clinical trial and operational datasets, there are huge hindrances to be overcome before this potential can be figured it out. In spite of the hindrances portrayed over, a few activities are in progress all over the world that are having some accomplishment in defeating these boundaries.

Informatics for Integrating Biology and the Bedside (I2B2), otherwise called I2B2, is an open source stage that incorporates growth patient's' electronic health records over a few Boston hospitals[19]. Researchers can think about the vast joined dataset to survey clinical results and recognize subjects for clinical trials. This will accelerate the tedious procedure of finding reasonable subjects and empower clinical trials to continue all the more quickly. I2B2 has now been additionally formed by associated doctor's facilities into an inquiry device named the Shared Health Information Network (SHRINE), which can mine the data and find up to this point unfamiliar zones of clinical learning. Startup companies are rising to utilize enormous data technologies to make more prominent utilization of clinical trial datasets right now held in different little possessions. Such new companies utilize progressively modern methods to consolidate datasets and extricate clinically valuable data from them. One such startup is Flatiron Corporation, backed by Google to the tune of $100m .The accessibility of joined data from clinical trials should help clinicians to recognize the most reasonable treatments for their patients, as they will have the capacity to perceive what has worked best with most patients in comparative conditions. They will likewise have the capacity to think about their own particular outcomes against different clinicians in comparative settings.

C. The Centre for Big Data Research in Health (CBDRH)

uses sophisticated algorithms to link large datasets containing information on hundreds of thousands of de- identified patients. By cross referencing and linking data drawn from a variety of sources, such as genetic profiling, social media, patient records, test results, claims data, home monitoring and mobile apps, researchers can identify patterns and trends that would not be apparent by studying a single data set in isolation

The UNSW electronic Practice Based Research Network (ePBRN) has prevailing with regards to joining datasets crosswise over hospital, pro facilities and essential care settings to support more prominent continuity of care for patients through a more incorporated health service delivery model.

X. WEARABLE DEVICES FOR BIG DATA APPLICATIONS IN HEALTHCARE

The presence of large amount of data provides an opportunity for technology to come up with solutions for key areas. One such area is the use of wearable devices to develop new tools for intelligent analysis and decision support tools for use in medical arena. Wearable Technology as an industry is growing leap and bounds, trying to meet the challenges of both healthcare professionals and patients. These devices are powerful, swanky and offer a variety of solutions for clinical use and patient related issues. It is also very important that users of wearable technology can rely on the data generation capability of the technology used. This technology makes it feasible to record vital statistics, data which doctors can later consult to make quicker and better informed diagnosis and choices of treatment.

The increasing popularity of wearable devices indicate a shift in perception by the population to look at preventive health care rather than corrective care. The wearable market worldwide grew by 29

% in 2016 according to the International Data Corporation(IDC). The popular units continue to be wristbands and watch like devices, though clothing and eyewear are gaining popularity. The virtual reality technology is one of the drivers for growth of wearable devices.

Fig 6: Wearable devices

A. few of top performing wearable devices based on a survey of available material in open domain are as follows:

1) The LIVE by Early Sense is a remote monitoring device that monitors sleep and vital statistics for bedridden patients. The piezoelectric sensor disk plugs into an outlet and slides under a patient's mattress. Real time patient vitals can be monitored through a mobile application, which can be supervised by caregivers and clinicians.

2) The Rapael Smart Glove has an exoskeleton design that assists people who have had strokes, and other patients with neurological and musculoskeletal injuries to regain mobility in their hands.

Fig 7.. Smart Glove

3) Rapael Smart Board , another product from Neo fect based on game based technology assists patients with shoulder and elbow mobility. [2]

Fig .8.Smartboard

of the classification available is simple devices with limited capabilities such as wrist bands , which may collect data regarding user sleep pattern, physical activity etc. Another type available is slightly advanced devices such as smart watches which has embedded OS which enables installation of third party apps similar to features available on a smartphone.

As mentioned earlier, the smart watches require an OS to run apps and services. Apple Watch which has almost 50 % market share is one of leading products available in the market. Using services available on platforms developed by the smart watch manufacturers like Apple, Samsung etc facilitative the users to gain access to their own collected data and third party developers

can build new applications which can use these collected data with certain strings attached. The *Apple Health* enables collection of data from apple devices , and it also supports some other companies such as Nike, Fit bit, Run Keeper etc. The Apple Health Profile has a general purpose profile and medical research profile , where the platform provides detailed data analysis for medical research. Similar other products are Google Fit+ Android Wear and S-Health (Samsung Platform supporting wearable's under its Tizen OS).

Variety of sensors fitted on these wearable device which collect data from the user . Say, the data collected in wearable can be transferred to a laptop or computer as an intermediate step prior to final transfer towards a permanent storage. This aspect is facilitated by a third party app generally. The Permanent data storage is provided by proprietary servers , where third parties and final users like say a medical research fellow can access the data.

Further, the data may be subjected to analysis typically a action performed in servers. The huge amount of data requires efficient methods of classification and analysis and it has been brought out that Deep Learning is a promising technique for large scale data analysis.

As with other aspects of healthcare technology, compliance to privacy and security remains one of the main concerns. Another challenge is the compatibility of these devices with software solutions, electronic health record, networks etc. Another important aspect which needs to be addressed is the change management among the medical fraternity to adopt this technology seamlessly.

B. Hypertension Scenario

For example, a patient or patient proxy is a pre- hypertensive and primary care provider(PCP) asked her to track her blood pressure(BP) twice a day. The patient then purchases a BP cuff and recorder BP in her daily log(technological or mechanical)[14].Each week, the patient then emails her readings via a secure email to her PCP. This is a hypertension scenario which highlights the flow of the process which involves data capture, data transfer and subsequently review/document. Data transfer refers to the communication of the captured data to the PCP or member of the health team by the patient by electronic or other methods whereas data review refers to the process of a PCP or health care member deciding what to do with the data after reviewing it.

C. An example of related work done

Hernandez et al. used a smart bio watch to monitor movement of a person using accelerometer and gyroscope sensors. Form these readings the Heart rate (HR) and Breath Rate (BR) has been calculated. To validate the results, two experiments were performed. First in a laboratory with 12 participants (male and female) having no history of cardiac and respiratory abnormality. second experiment was conducted with three people wearing these watches during sleeping. From both of these experiments the calculated values of HR and BR were transferred to the database and compare the performance in different combination.

XI. SECURITY OF BIG DATA IN HEALTHCARE

The huge advances in handling and interpreting voluminous data being generated has made Big Data technology a driver for growth in medical

Field. However, privacy and security concerns in handling medical data are foremost and needs to be addressed. he healthcare system is generally distributed system where data needs to be accessed from various locations and Big Data solution for the healthcare should provide privacy and security. The public needs to educate on the potential value of big data in the field of healthcare. To adopt the technology effectively, total commitment from high level government and medical community with consistent investment over a long term are the key drivers .

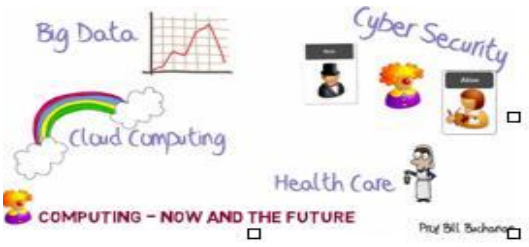


Fig 9. Future of Computing

XII. DATA VISUALIZATION

Privacy is aspect which is very dear to individuals and needs to be respected. Technology cannot hold us slave and permit intrusion into our personal and private life. Our medical data is very personal and extreme safeguards have to be put in place to make sure that the information only gets to whoever is meant to see it. Just imagine getting a call from a renowned hospital offering to treat your nascent disease at a low cost compared to Hospital X …it is frightening. Despite the best practices, cyber thieves routinely target medical records. It is said that they earn more money in pilferage of medical data than from theft of credit card data. In Feb 2015 , Anthem, the second largest US health insurer reported loss of medical data related to 80 million patients. Fortunately, names and addresses were hacked not illness data. Some experts have called for an International Regulatory body to oversee security of data in healthcare.

The reference highlights a methodology for transmission of medical data over a distance and online where combination of image cryptography, data hiding and Steganography technique for de- noised and safe image transmission purpose was used. Estonia in EU has made data security as a major focus area by investing after it learned the hard way post a cyber-attack in 2007. The medical data is stored in encrypted form thereby providing security. The hospitals also have to invest in their IT infrastructure to provide robust and secure systems. Use of block chain network for hosting medical data has also been practiced by various organizations. Block chain network technology which was used for Bitcoin has phenomenal applications in other fields and provides excellent security overlay. This can be effectively used in data security for healthcare.

For efficient decision making in the healthcare industry, it is imperative that we not only communicate insights to the primary health care providers in manner which yields the maximum results but we need to create data visualization tools so that healthcare providers can relay information to the patients as well[20].The utilization of visualization inside the healthcare domain has a long and storied history. In any case, there are new and one of a kind challenges rising in the present data rich healthcare services industry where modern interactive visualization techniques can assume a critical part. The tremendous capability of these methods is reflected in a few late advancements. For example, recent articles inside the visualization literature have given surveys of rising exploration focusing on special healthcare related research issues.

Data related issues in medicinal services are comparative from multiple points of view to those in different spaces.[20] Challenges of data integration, wrangling, usability, and interpretability are largely focal issues. Be that as it may, the healthcare discipline additionally presents various area specific challenges:

A. Expansiveness of utilization, from individualized point of-care to vast scale populace healthcare applications;

B. Data's multifaceted nature, including vast quantities of patients, extensive quantities of heterogeneous variables, information connecting over numerous sources, and absent or fragmented information; an

C. Statistical r i g o u r , where "intriguing or interesting" isn't sufficient given the desperate stakes inside the medicinal services space.

By tending to these difficulties and incorporating the current workflows of healthcare specialists, intelligent information driven visualization can possibly turn into a valuable, and maybe fundamental, apparatus for a cutting edge data driven healthcare system

XIII. CONCLUSION

Big Data is very important for making personalized medicine a reality. As high level technology cannot happen in isolation, health data has a limited impact if stored and used within an isolated computer by one doctor. The real benefits come from data sharing and access to data, with full respect of all data protection mechanisms. This approach will enable the technology to be used gainfully for betterment of quality of human life. Data harmonization and integration is the key in providing single patient view for better diagnosis and treatment to deliver quality health service.. This enables the healthcare providers to also bridge the gap between the insights that the data provides and the implementation of this data analysis. Patients and healthcare professional both profit from Big Data in various ways. Further practical approaches are needed to adequately balance the benefit and threats of more and more detailed and sensitive data being available. There is a need for technologies that can handle multi-modal data in an efficient manner. At the point when joined with contemporary business knowledge capacity, the investigation of consolidated datasets from intense care (hospital) services, specialist clinics and primary care service (General Practice) results will empower clinicians to: All the more precisely foresee results that guide treatment choices Produce new information about the effectiveness of treatments Enhance compliance with rules;

Enhance rates of return for follow up care; Enhance, energize healthier way of life, wellbeing; Energize patient involvement.

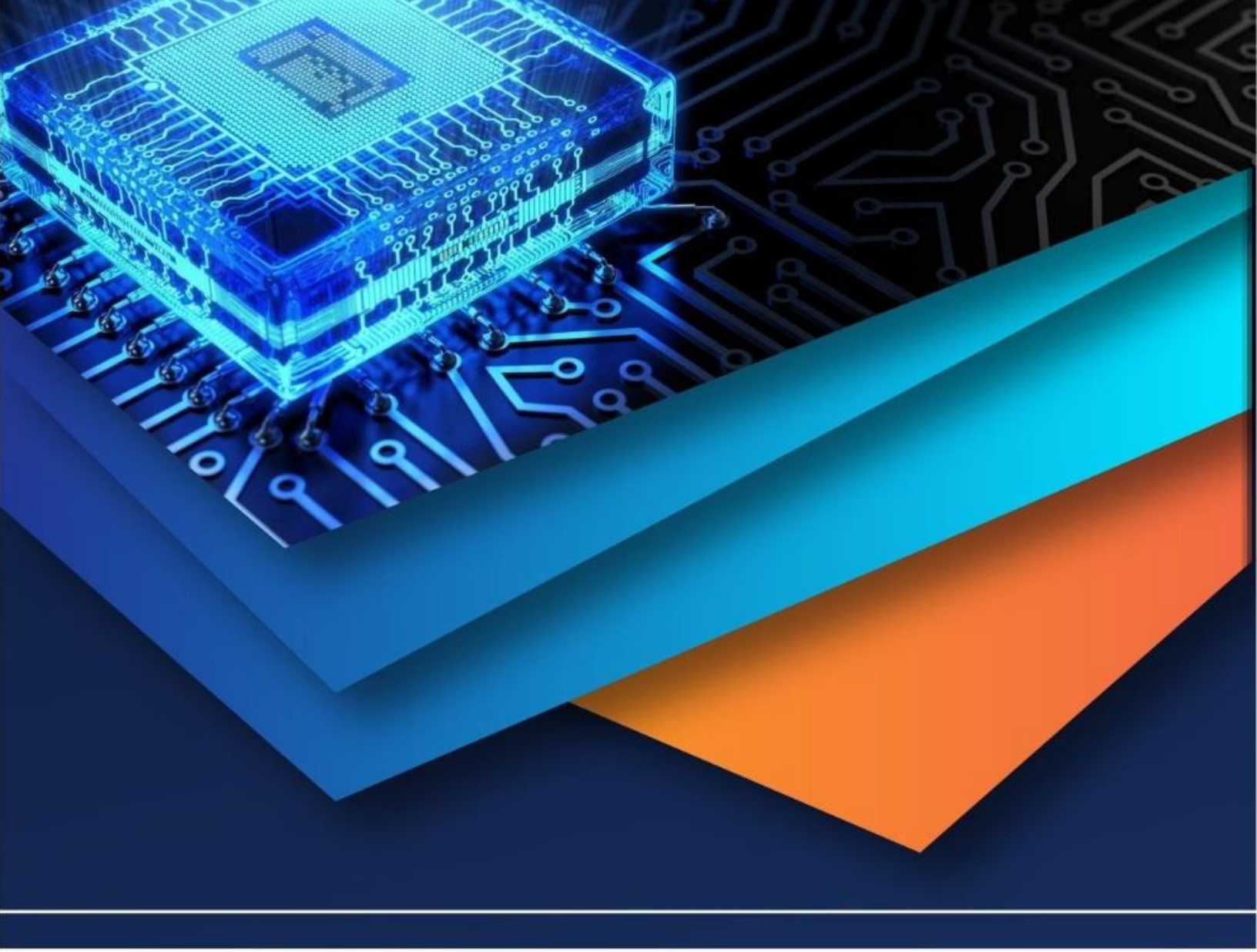
XIV. FUTURE WORK

Most of the deaths are due to major chronic diseases. In this paper we focuses on heart patients that are increasing rapidly day by day. To improve
the health quality of these patients we propose a complete big data e-health framework. For data collection we proposes an electronic health band with multiple sensors, for preprocessing we will use multiple algorithms to clean and merge the data, for analysis we use some machine/deep learning algorithm and for data presentation we develop some dashboards for doctors and some mobile applications for patients.
The next step is to implement this proposed big data frame work. We will collect the data from register patients of our selected hospitals. We also train patients and doctors to use their applications [22].

REFERENCES

[1] Why Healthcare May finally Ready for Big Data Nilay D.Shah & Jyotishman Pathak, Harvard Business Review
[2] Top 10 healthcare Wearable to Watch DonnaMarbury Marc 2017
[3] Wearable technologies – future challenges for implementation in healthcare services HadasLewy, Feb 2015
[4] Francisco de Arribaperz et al, " Collection and Processing of Data from Wrist Wearable Devices in Heterogeneous and Multiple-User Scenarios" ,2016
[5] Dimiter V. Dimitrov, "Medical Internet ofThings and Big Data in Healthcare"
[6] Ping Jiang, Jonathan Winkley, Can Zhao, Robert Munnoch, Geyong Min, and Laurence T. Yang, Member, "An Intelligent Information Forwarder for Healthcare Big Data Systems With Distributed Wearable Sensors
[7] Dhiya Al-Jumeily, Abir Hussain, Conor Mallucci, Carol Oliver, "Applied Computing in Medicine and Health"
[8] Bernard Marr, https://www.forbes.com, "How Big Data Is Changing Healthcare"
[9] Bikash Kanti Sarka, "Big data for secure healthcare system: a conceptual design"
[10] Steve Bridges, euobserver, "Digital health has potential but needs to be secure"
[11] TF7 Healthcare subgroup, BDV association,"Big data technology in healthcare"
[12] Shahidul Islam Khan, "Health Data Integration with Secured Record Linkage"
[13] Philip Dabek, "Leveraging Big data to provide a web service that provides the likelihood of developing psychological conditions after a concussion"
[14] Michael Shapiro, Douglas Johnston, Jonathan Wald, Donald Mon, "Patient generated health data", RTI international 2012
[15] Emanuel Wietschek, Giovanni Felici, Paola Bertolazzi, "Clinical data mining",2013
[16] Dr. Aynur Unal, "Security in Big data of Medical Records"
[17] Reena Duggal, Sunil Kumar Khatri,Balwinder Shukla, "Improving Patient Matching: Single patient view for Clinical Decision support using Big data Analytics",2015
[18] Dr. W Liu, Dr. E.K Park, "Big data as an E- Health service"
[19] John Lewis,Prof siaw-Tieng-Liaw, Pradeep Ray, "Applying Big data and business intelligence insights to improving clinical care for cancer",2015 IEEE
[20] David Gotz and David Borland, "Data Driven Healthcare:Challenges and Opportunities for Interactive Visualization",May/June 2016 IEEE Computer Society
[21] Raghunath Nambiar, Ruchi Bhardwaj, Adhiraj Sethi, Rajesh Varghese, "A look at challenges and opportunities of Big data analytics in healthcare",2013 IEEE
[22] Atif Naseer,Basem Y. Alkazemi,Ehsan Ullah Waraich, "A Big Data Approach for Proactive Healthcare Monitoring of Chronic Patients" 2016IEEE
[23] Mary Jo Deering, "Patient-Generated Health Data and Health IT" 2013 The Office of The National Coordinator for Health Information Technology,USA.
[24] C Imthyaz Sheriff, Twaseef Naqishbandi, Angelina Geetha "Healthcare Informatics and Analytics Fra

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue XI November 2017- Available at www.ijraset.com



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY