Atulyam Bharat Is Not Possible Without Swastha Bharat – Digging Down With Our Big Data Challenges

Shalu Gupta[1], Prof. (Dr.) Ganesh Gopal Varshney[2], Prof. (Dr.) Pooja Tripathi [3]

Research Scholar,Mewar University,Ghaziabad, India[1]

Mewar University,Ghaziabad, India[2]

Inderprastha Engg college,Ghaziabad, India[3]

Abstract: 'Atulyam Bharat' or we can say Incredible India is the dream of every Indian including our own Prime Minister Mr. Narendra Damodardas Modi. But this dream can never be achieved without the 'Swastha Bharat' or Healthy India. Health and well-being of its citizens is the first priority of any country and so as India. No one can deny the fact that India seeks to become a world power. Better population health is one among the vital considerations in India. All those who are living in cities and massive cities have no doubt access to high finish health services, the ample people living in rural India, notably within the remote elements of the country face issues of inadequate facilities and poor access to attention. The inefficiencies and inequities within the public health care access in India have pushed forward the necessity for power and innovative solutions to strengthen a similar. Paper identifies the large deficiency and scarcity of right and timely health care facilities to our Rural India. We also point the addresses a way to give bigger access to primary health care services in rural India. In this paper, it conjointly addresses the important computing and analytical ability of Big Data in process huge volumes of transactional information in real time things to show the dream of Swastha Bharat (Healthy India) into reality. We have proposed a model to predict the diseases based on the various symptoms provided by experts.

Keywords: Big Data Analytics, Health care in India, Big Data Challenges, Big Data, Rural Health Care, e-Health Care, Swastha Bharat, Atulyam Bharat.

I.   INTRODUCTION

"To keep the body in good health is a duty, otherwise, we shall not be able to keep our mind strong and clear." - **Buddha**

The healthier you keep your body, the stronger, the clearer, and the better your mind will function. Hence, make sure that keeping a healthy body is the most important duty of yours. A beautiful health quote. If we have a tendency to reach achieving this revolution, people expenditure on medical treatment are significantly reduced, and then they'll use their savings for higher things, he remarked.

As day-by-day the population is increasing in Asian countries, it has over-burdened the health care infrastructure within the country. Health care in our country is government supported and government run. But, for several individuals living within the remote components of the country, accessing primary health care continues to be a challenge. For the developing countries like India, the health care is one among the foremost necessary areas to stress, therefore on give higher health care access to the valuable human resources, that successively will build the India healthier too. In India, massive amount of data has been generated by the healthcare industry traditionally and historically, driven by keeping these records, compliance & restrictive needs, as well as patient care taking and various other services. These types of data when managed electronically is known as Heath Data Sets. As they are massive in nature and complicated too that they are difficult to manage by already existing code as well as information management tools and ways. All the scientist related to Big Data domain there is a big challenge to find out the associations and fruitful patterns among the huge bunch of data is a big challenge. The analysis used in big data applications of health care has boomed up the market by extracting the knowledgeable information and decision making to boost health care, health insurance issues, save lives, and lower the prices.

Looking into the matter seriously, the digital health care solutions has reworked the health care system to become far more economical, more cost-effective and bring home the bacon higher quality than before with the introduction of big data analytics in this health care sector that build the performance of health centre higher, and folks live healthier.

The following area unit the key points that build the performance of health centre higher and folks live healthier.

1. Measure, store and analyze the data to boost treatment quality
2. Manage revenue prices by reducing unnecessary tests
3. Improve preventive care and increase patient & physicians' satisfaction

Big Data Analysis has a great potential to take out better results. Whether it is through analyzing the patient characteristics, analyzing the price tags to provide better and value efficient medical and health care facilities.

II. HIT IN RURAL HEALTHCARE

The known Indian economic expert and Nobel Laureate man "Amartya Kumar Sen" has mentioned that "Growth in value by itself isn't enough, if the advantages don't manifest themselves within the variety of additional food, higher access to health and education".

Health information technology (HIT) has the power to boost the standard, safety, effectiveness, and delivery of attention services in rural communities. HIT will connect rural patients and suppliers in distant locations to specialists in urban areas.

HIT uses technology to store, secure, retrieve, and transfer protected health info electronically among attention systems and community settings.

- Electronic health records (EHR) for patients, rather than paper records
- Secure digital networks to send and deliver up-to-date records whenever and where the patient or practician may have them.
- Electronic transmittal of medical test results.
- Telehealth applications to extend access to suppliers.
- Confidential and secure patient health portals for patients to access their personal health info on-line.
- Transmission between attention suppliers, additionally as patients.
- Electronic prescribing and ordering to assist avoid medical errors.
- Call support systems to produce clinicians with info on best practices and treatment choices to boost quality of care.
- Mobile devices and tablets to update patient records in real time and document at the purpose of care.

The impact of those IT/ICT tools on the lifetime of natives world over has been largely witnessed by the urban population. The rural population virtually world over has been the smallest amount beneficiary of such technological developments with the result that the digital divide between the agricultural and therefore the urban still remains an area of concern. It is in this setting an endeavor has been made by the creators to evaluate the invasion of Information and Communication Technologies in Present Health Care System in India. Altogether, we can say that these increases in data liquidity have had bring the whole industry to the tipping point.

Dr M.C. Misra, director of All Republic of India Institute of Medical Sciences (AIIMS), has highlighted during a speak that "Advances in medical technology and new medicines square measure so a boon, however, to figure in Republic of India they need to be price for cash. the general public can't even afford typical treatments at backed costs publically hospitals".

A study given within the November 2013 issue from Harvard Business Review that due to low on price and high on quality of health care, the sector not only has attracted over 1.3 million patients from abroad annually but with a large vary of treatments out there — the Indian heath care system. It has generated $3 million by the top of 2013. The authors Vijay Govindarajan and Ravi Ramamurti also gave an affirmation to private hospitals in India for "delivering foremost and world class health care, affordably".

The World Bank data illustrates that although we are providing world class care to the world but, 99% of India's population cannot manage to procure these services. They have also extended that every year, thirty-nine million individuals square measure pushed into poorness by defrayment out of their pocket for tending, with households on the average provide over five. 8% of their expenditures to treatment, the info exposes.

To access the high-end non-public sector facilities, individuals square measure payment giant amounts of cash on treatment that in-turn affects their livelihoods and slowly poor individuals are getting poorer. In such cases, the idea of Telemedicine along with big data analytics is that the higher various to treat the patients in rural areas.

III.   BIG DATA IN GENERAL

Definition: Big Data could be a assortment of enormous and sophisticated data sets that unit of measurement hard to technique victimization common direction tools or ancient process applications. "Big Data refers to the tools, processes and procedures allowing a corporation to form, manipulate and manage very large data, information and knowledge sets and storage facilities".

Since the digitization started with in very less time the Big Data has produced by each and every sector in India, every social media and digital technique exchange produces it. Systems, sensors and mobile devices transmit it. Big Data is obtaining back from multiple sources at associate degree fearful speed, volume and choice. To extract vital value from large information, you'd like optimum method analytics capabilities, power and skills. Big data can be called as go-ahead the approach people at intervals organizations work on. The working culture within the business is changing and the decision makers and IT professionals ought to get along to understand the value from all information. Higher decisions - optimizing operations, preventing threats and fraud, capitalizing on new sources of revenue and deepening consumer engagement are the major achievements we can get from large data. A new approach to style, tools and practices is required to escalating demand for insight. The large data area unit generated from on-line transactions, search queries, health records, science information, emails, videos, social networking interactions, sensors, audios, images, click streams, logs, posts, mobile phones and their applications

3.1 Various Characteristics of Big Data

Volume
The total volume or amount of data generated by different departments of health care sector now-a-days seems to be larger in size as compared to previously existing data. TB to XB

Velocity
The frequency of this total amount of data generated by different departments of health care sector are increasing day-by-day as the number of equipments are increasing in the health care sector.

Volume	Velocity	Variety	Veracity
Data at Rest	Data in motion	Data in many forms	Data in Doubt
TB to XB of data to process	Streaming data	Structured, unstructured	Uncertainty due to data ambiguity

Figure 1 Big Data – 4 "Vs"

Variety
The data already existing in the heath care sectors square measure less to adjust new quite detector data kinds, activity etc. Whereas non-traditional data like medical equipment's square measure merely adaptable to change.

Veracity
Bid data is collected through vast sources. Veracity deals with imprecise or unsure data. There was forever the assumption in ancient data warehouses that the knowledge is precise, definite, and clean. But it isn't therefore among the case of massive knowledge.

3.2 Advantages of Big Data to Health Care

Spending, Prices, Costs

• "Spending" refers to what is spent on health care commonly. Thus, if people use plenty of health care services, this causes payment to increase even supposing prices keep identical.

• "Prices" see what health care suppliers charge for express services and merchandise.

• "Costs" see what health care suppliers pay therefore on turn out express health care services and merchandise to patients. This is often resembling prices minus profits or losses.

IV. WAYS HOW BIG DATA IS IMPROVING HEALTH CARE ANALYTICS

There are some ways through that the fix rural health care in India and bridge the gap between quality and affordability in government hospitals. These ideas can change North American country to access the services on par with the non-public super speciality hospitals. Further, the implementation of those ideas can give cheaper, higher and easier health care facilities to the rural people of India.

1. **e-Health File:** The creation of a e-Health care file for every patient, wherever all health care suppliers and patients themselves were able to submit data (with the consent of the patient) each subjective knowledge, symptom diaries, lab data, image medicine, pathology reports etc, may be filed to beat the data overload from the large amounts of knowledge. Analytics in big data may be used for the process of the data and acquire the required results with nice accuracy in cheap time.

2. **Making awareness with chronic diseases:** The system should establish and make awareness among the folks with the common chronic diseases at specific areas, through that we will forestall diseases. These chronic diseases are chargeable for the seventy fifth percent of health care disbursal owing to lack of awareness and prior care.

3. **e-Prescribe:** Paper primarily based prescriptions are archaic and result in many miseries every year owing to errors in prescription. However if each doctor is supplied with a modern electronic prescription system which a suggested as a new way for prescription, it would definitely improve safety of the patient by creating prescriptions easier to browse and providing instant checks on dosages, drug interactions and a patient's medication history.

4. **e-Medical Records:** Doctors agree that e-medical or the electronic medical records are a requirement for the higher health care in India. But, at the present solely few hospital are maintaining EMR's, chiefly as a result of price, privacy problems, and also the lack of the compatible and of course user-friendly infrastructure.

5. **Stop needless Treatments:** Doctors ought to avoid trial and error variety of medication. The matter should be examined totally by acting the specified diagnostic tests throughout the preliminary days of sickness. The proper treatment ought to be prompt at the primary visit solely that avoids the sickness to become a lot of essential. Most of the issues are arising with the misdiagnosis and wrong treatment throughout the first stages.

6. **Scale Back Fatality Rate:** The baby Mortality rate in our country is well massive compared to different countries. Though, the government has many schemes for the pregnant ladies, those weren't yielding higher results owing to the shortage of correct treatment. If the correct care is taken towards the pregnant ladies, positively the new-born baby are healthy all told aspects to create India healthy.

7. **Tele Medicine**: Doctors will typically diagnose or impose while not seeing the patient. The patient needs to physically seem before the near sickbay, wherever the nurses or medical examiners can diagnose initially level, note the symptoms and informs the high level specialist doctors regarding the case. When examining the reports, the specialist doctor suggests the treatment through medical examiner that reduces prices and creates satisfaction by virtual communication of doctors with patients to debate any changes in medication if required and check the results also throughout a web system.

V. HADOOP INTRODUCTION

A Hadoop is a programming framework which is Java-based. It is an element of the Apache project sponsored by the Apache software system Foundation which supports the process of huge knowledge sets in a very distributed computing setting. Hadoop was originally planned on the premise of Google's Map cut back, during which Associate in Nursing application is attenuated into various tiny elements. The Apache Hadoop software system library will observe and handle failures at the appliance layer. The Hadoop in the main includes:

1) Hadoop Distributed classification system (HDFS)

2) Hadoop Map cut back

VI.   BIG DATA ANALYTICS IN HEALTHCARE THROUGH HADOOP

As far as data analysis is concerned in heath care hadoop has primarily changed the scientific discipline of storing the information and then analyzing it. An ascendible electronic information service worth $100K per TB was in use since 5 years. For providing maintenance and support for the same one should have a perpetual package license of $20K per annum.

But now a days with a $1,200/year subscription we did able to manage store and analyse an identical amount of information.

These new distinct features and reduced prices has attracted the scientist society and have had make Hadoop as a center of attraction for mainly large-scale data management activities and analyses can either originate or integrate. Further with durable SQL capabilities the Hadoop infrastructure has been coupled and transfer the complete SQL-based subject to the platform of Hadoop. The market for Hadoop has expanded by one or a pair of orders of degree. Also Hadoop is not remained merely as the domain of specialists only every data warehousing organization are expanding to it.
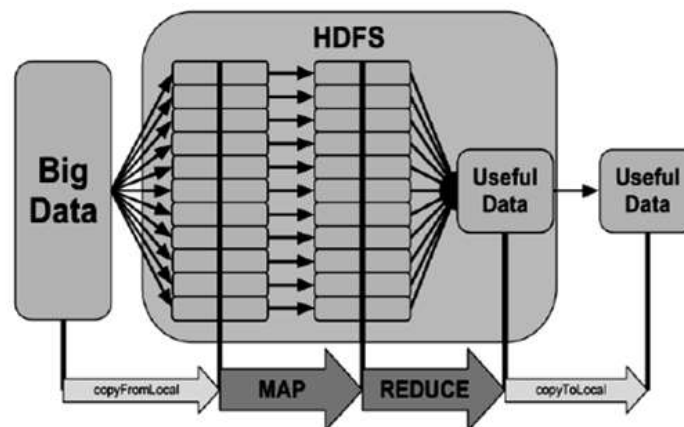

Figure 2 Hadoop's Map Reducing

Relational technologies in operation on their own face area unit another obstacle. The jinni is out of the bottle with relevancy to the traditional multi-structured, loosely-structured and unstructured data. With Hadoop we have a tendency to currently analyze a lot of and totally different info than we are able to with relative databases. Across economical and high level servers distributed data processing may be expedited on giant volume of data by Hadoop, it may be scaled on the far side limits and used for storing and processing the data. No data is also thought of for Hadoop as excessively massive. Within the current speedily growing medical world huge live of knowledge is being made and added daily. Thanks to Hadoop's potency and effectiveness the data thought of earlier well for the analysis currently loses their price. A distributed filing system is used by Hadoop for storing files on numerous systems throughout the cluster. Hadoop has also a great potency in hiding the files location within the cluster being accessed the top users will reference files identical approach they are doing it within the native system.

VII.   PROPOSED ARCHITECTURE

Our proposed interactive healthcare system architecture has focused on three phases which were shown in Figure 3 in the first layer we are collecting the data with the unique sensing interaction of smart biosensor devices, then-after analyses will be done on the massive volume of data collected by individual patients as well as hospitals for future disease prediction, future reports and find the pattern of similar patients.
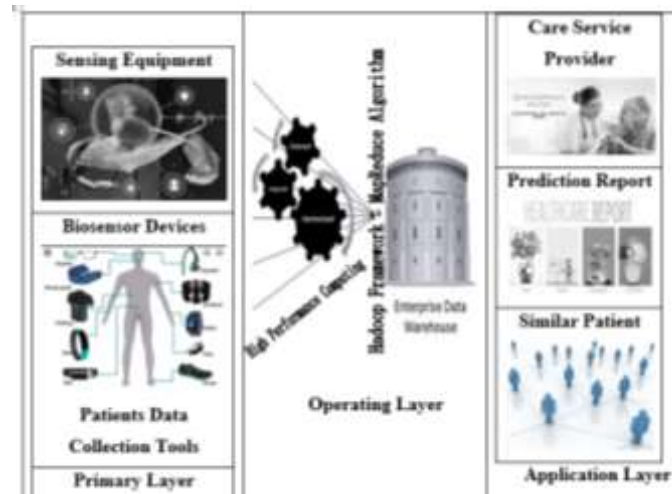
Figure 3 Three phases of interactive healthcare - Primary layer, Operating layer, and Application Layer

These three phases are executed in three different layer called primary layer, operating layer and application layer. In our research, we will discuss the architecture of those three layers and formulate the architecture of the cloud-based server for the system.

VIII. PROPOSED ALGORITHM

Predictive model is a created function of input feature or symptoms of a disease and the output is the predicted disease which can be mathematically represented in Equation-4.7 as

$$T = F(\psi_f) + e$$

Where T is the predicted target disease, ψf is the set of target feature which are extract

Algorithm 2 Genetic algorithm

INPUT: Dataset = D
Feature = f
initial-population = δ
Iteration = 100

OUTPUT: Fitness value of feature.
1.　　　Set crossover probability = 0.8.
2.　　　Set mutation probability = 0.1.
3.　　　Construct model, model = linear-regrassion(T, D).
4.　　　x = matrix(model).
5.　　　y = response(model).
6.　　　while Iteration != 100 do
7.　　　δ = crossover(δ-1)
8.　　　δ = mutation(δ)
9.　　　f = ComputeFitness(δ)
10.　　　end while
11.　　　return f = best feature
12.　　　<u>Fitness function</u>
13.　　　ComputeFitness(input String)
14.　　　Set Random Fitness = where(String = 1)
15.　　　X = combine(1,x,[,Fitness])
16.　　　best-fitness = linear-regrassion.fit(X,y)

IX. PROPOSED MODEL

During supervision and learning we are training our Healthcare prediction model by analyzing existing data because we already know whether each patient has heart disease. The trained model is then used to predict if users suffer from heart disease. We incorporated genetic algorithm to eliminate the irrelevant feature. The fitness of feature over a generation of heart patient dataset. In genetic algorithm, we use the default mutation and crossover probability which is 0.1 and 0.8

respectively. After extracting features, we apply different classifier such as **Logistics Regression Technique (LR), ANN, Classification Tree etc** to predict a patient's disease. We also apply those classifiers before we implement the genetic algorithm to compare the accuracy and performance. Accuracy Parameter during Training using LR Technique comes out as 89% which is best.

ROC Curves summarizes the trade-off between the true positive rate and the false positive rate for the predictive model using different probability thresholds. The accuracy of the ROC curve came out to be 83.244%
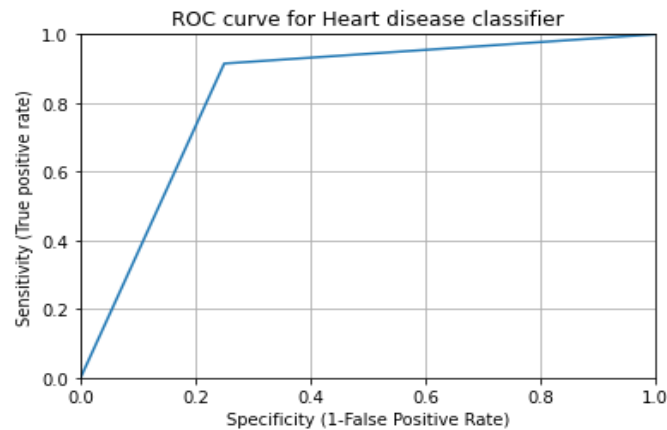

Figure 4   ROC Curve for Heart Patient Dataset

```
import sklearn
sklearn.metrics.roc_auc_score(y_test,pred)
```

```
0.8324468085106383
```

Figure 5   The accuracy of the ROC Curve

X.   CONCLUSION

In this paper discusses the Big data analytics in the field of heath care is growing into a capable field for given that insight from terribly data sets and outcomes whereas reducing prices. The potential of big data using Hadoop is immense, but there stay challenges to beat. Big Data analytics has the capability to remodel the manner aid suppliers use refined technologies to achieve insight from their clinical and different information repositories and create wise selections.

In the health care organization, the implementation and use of Big Data Analytics will widespread very fast in near future. The decisions based on this analysis and use of the final information came out from it will aid the health care organizations.

This paper further discusses regarding the big data and its characteristics, strategies and challenges and suggests the way to overcome the underlying issues being round-faced by the health care trade. Additionally, presents the massive or big ideas to repair the healthcare system in India. The implementation a part of this paper may be done victimization HDFS (Hadoop File System) for the huge information storage and Hadoop Map cut back with Amazon internet Services. The employment of big data analytics across the health care organization and care trade can mine the doctor's research lab transcript's victimization text mining and correlation to patient outcomes and placement aware application analytics for enhancing client expertise. Achieving better outcomes at lower prices has become vital for health care which might be achieved through the implementation of this paper victimization Hadoop HDFS and MapReduce to uncover the data lying in big health data sets.

This research describes a new healthcare system for a **'Swastha Bharat',** which can deal with the large volume of patient's data along with the full overview of patient's medical conditions. Besides data collection, we have designed a probabilistic data acquisition scheme which will help to analyze the massive amount of unstructured data and those schemes are efficient for a loud environment. After a regressive study we have suggested to enable a two-way interaction with HPC and cloud server by introducing a data warehouse to store data and empowers other functions. In this research, we have also implemented some prediction model algorithm on existing dataset and showed the performance of those models. We have showed a comparison of Logistics Regression (LR) data mining technique on healthcare data.

XI. REFERENCES

[1]   Jeremy Shiffman, Bloomberg, The Role of National Health Information Systems in the Response to COVID-19, Bloomberg Distinguished Professor of Global Health Policy, 2020

[2]   Jiang, P. ; Winkley, J. ; Zhao, C. ; Munnoch, R. ; Min, G. ; Yang, L.T. Systems An Intelligent Information   Forwarder for Healthcare Big Data Systems With   Distributed Wearable Sensors, IEEE, Journal

[3]   Zolfaghar, K. ; Meadem, N. ; Teredesai, A. ; Roy,   S.B. ; Si-Chi   Chin ;Muckian,   B. Big Data, 2013 IEEE International Conference , Big   data solutions for predicting risk-of-   readmission for congestive heart failure patients

[4]   Valentine E. Balas, Raghvendra Kumar, Rajshree Srivastava, Recent Trends and Advances in Artificial Intelligence and Internet of Things, Pages 644 [618] Year 2019

[5]   D. Peter Augustine, Leveraging Big Data Analytics and   Hadoop in Developing India's   Healthcare   Services, International Journal of Computer   Applications (0975 – 8887)Volume   89 – No   16, March 2014

[6]   El-Gayar,O. ; Timsina,P. System Sciences (HICSS),   2014 47th Hawaii International   Conference in   IEEE   Opportunities for Business Intelligence and Big   Data Analytics   in   Evidence Based Medicine

[7]   Nambiar, R. ; Cisco Syst., Inc., San Jose, CA, USA   ; Bhardwaj, R. ; Sethi, A. ; Vargheese,   R. , A look   at   challenges and opportunities of Big Data analytics in   healthcare, IEEE(2013)

[8]   Zolfaghar, K. ; Meadem, N. ; Teredesai, A. ; Roy,   S.B. ; Si-ChiChin ; Muckian, B. BigData,2013IEEE   International Conference , Big data solutions for predicting risk-of-   readmission for congestive   heart failure patients

[9]   Srinivasan, U. ; Arunasalam, B. Leveraging Big Data Analytics to Reduce Healthcare Costs,   IEEE   Journal

[10]   Patil,HarshKupwade ;Seshadri,Ravi Big Data (BigData   Congress), 2014 IEEE International   Congress   on   Big Data Security and Privacy Issues in   Healthcare