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

A Survey on Disease Prediction by Machine Learning Using Big Data Analytics

Syed Ahmed [1], Ashish Savant[2], Gaurav Ghodke[3], Aishwarya Kamble[4], Dnyaneshwar Bhise[5]
[1]*Research Scholar, KLU & Assistant Professor, JSPM JSCOE Pune.*
[2, 3, 4, 5] *Computer Department, JSPM, JSCOE Pune Savitribai Phule Pune University*

Abstract: Big data is collection of large amount of data. This is weigh big. Which is uncontrollable by the usual method It is difficult to process this large amount of Data in traditional ways. So there are many techniques to handle and analyze this large amount of data. The challenge we face while storing this huge amount of data is analysis, sharing, storage etc. Big Data is difficult to work with the traditional approach hence there are different methodologies. The different form to handle with term is now applicable to the different fields industries. This survey paper describes the techniques used by researchers for predication of disease using machine learning, A.I techniques and big data.
Keywords:- Big data, Healthcare, Machine Learning, Big data analysis Management.

I. INTRODUCTION

A. What is Big Data?

The large amount of data producing daily over entire globe.This much of data coming from each and every where. The data may be in the form of videos, audios, social media, digital pictures etc. This all together known as big data it is difficult process with traditional methods and software techniques. Big data have the potential to improve records and make it faster. Big Data is now become common in IT industries. Big amount of large large data is available in industry, according to this data is divided into varies formats for analysis. Consisting records of millions or billions of users all from different sources. This basically loosely bound storage of data[10].

V's in Big Data are as follows:-

1) *Volume*- Volume in big data termed as the Size of collected data over entire plane from Different sources. There are many factors Contributing the volume of big data.
2) *Variety*- Variety declaring the different types and formats of big data. Data today comes In different formats with structured and Unstructured etc.
3) *Velocity*- Data is streamed with the unpredictable velocity which is cannot mapped the reaction time for such that calculated with velocity with which it is retrieved or uploaded.
4) *Variability*- In the flow of variety and velocity the next important factor is variability which indicates the variable nature of data. It differ Down the data from its main source stream
5) *Veracity*- It encompasses the other V"s. edacity provides the confidence in the Truthfulness of data.

B. Advantages of Big Data

1) Competitive advantages of big data is in business, healthcare industries etc.
2) Big data is large amount of chunks of raw data that is collected, stored, and analyzed through various means which can be utilized by organizations to increase their efficiency and take better decisions
3) Big data is more secure, relevant, Holistic, trustworthy, timely, accessible.
4) .It improves the performance in huge manner so as to become faster in processes Processing increase in rapid way with it use.

C. Why Big Data is Different?

In the old days a few years ago, we should Utilize systems to extract, transform and load data (ETL) into giant data warehouses that had business intelligence solutions built over them for reporting, periodically, all the the systems would backup and combine a data the systems would backup and combine a data everyone gets insight into what was going on. The problem was that the database technology Simply couldn't handle the multiple continues Stream of data. It couldn't handle the volume Of data. Big data solution offer cloud hosting Highly indexed and optimized

data structures. Automatic archival and extraction capabilities And reporting interfaces have been designed To provide more accurate analyses that enable Businesses and in healthcare to make better Decisions. Because the availability of large Amount of data in real time. Collection and Processing of huge volumes provides the Ultimate results [10].

Fig: 1 Benefits of Big Data

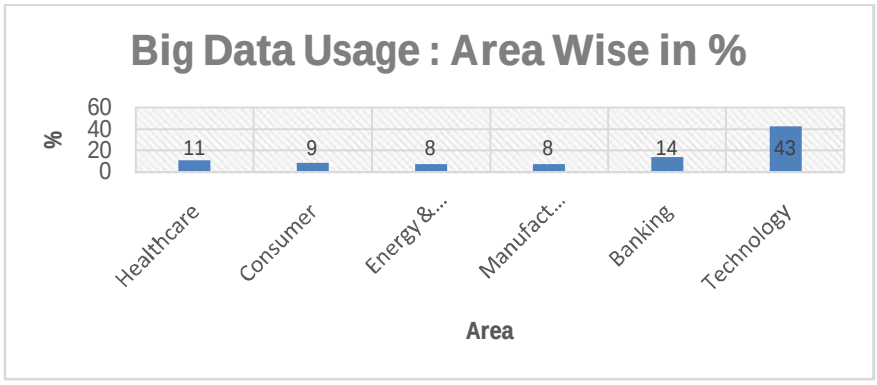
When the overall analysis of big data is under consideration then effectiveness and benefits are the two most important aspects taken under considerations. Cost reduction is a key factor which becomes handy for users. The decrease in cost helps for further processing with the help of big data. Decision making process becomes easier and faster which enhances the performance. New products and services be able to create with the maximum use big data.

D. Issues in Big data

1) Big data sometimes violates the privacy and the privacy principle.
2) Data can be used for manipulating customers. This is seen from various fields.
3) Big data may increase social stratification. Which imbalances the social proportions
4) Faces difficulties in parsing and Interpreting. Sometimes this thing happen which slows down
5) the system.
6) Big data is difficult to handle, as we See that large amount of data is generated on daily basis.

E. Big Data in Healthcare Analysis

Healthcare database is available into large amount data over globe. This data Is in the form of hundreds of terabytes, Petabytes. Healthcare data is divided into Different subfields of data according to Patients diseases and treatments on going. To evaluate and make out the necessary Data out of it requires different algorithms. Techniques like Map-Reduce is one of the Most efficient way retrieving and processing Of such huge amount data. Healthcare data Stored according to patient. Management of This data is very important and necessary. Innovation is key aspect in this field. So as to Make it more faster for analyzing, processing, and retrieval of healthcare data[5].



Graph 1: Area Wise Big Data Uses

According to analysis on uses of big data is in all the relative fields. The overall uses of big data in each and every field is varying constantly. Most of the use or utilization of big data is in technology sector where growth is rapid as new researches coming in. In below that healthcare, banking, manufacturing, consumer, energy and other sectors uses big data for enhancement and better

performance. As we seen banking and healthcare data sectors having uprights for evolution of big data. The use of big data will surely increasing in those areas in future.

F. Technical perspective in Healthcare Data

Health care data available in the heterogeneous Formats. Tools for the technical perspective Requires the input database which contains Appropriate numbers of patient records. We Applying here Map-Reduce technique to form cluster of data. This data mapped with the help of mapper. It maps the disease related data with respect to patient. Reducer reduces or breaks down the data into chunks. This technique improvises the entire process[5].

II. LITERATURE SURVEY

Table: 1 literature Survey

Paper no.	Techniques/Methods/ Algorithms	Tools	Journal	Year
1	Data mining classification algorithms, DMED & k-NN classifier , Dynamic Map Reduce framework.	WBANs , EHR,DICOM	IEEE	2017
2	Map Reduce Encryption/ Decryption , G-Hadoop Framework.	Hadoop , HDFS	IEEE	2016
3	Various Outlier Detection Techniques - Spatial , Classic Statistical Approach, Density Based Approach, Space Based Approach, Graph Based Approach.	Credit Card, Calling Cards	IEEE	2016
4	HITS Algorithm, Healthcare record sequence diagnosis, Healthcare frequent pattern recognition, Anomaly detection with RTM.	Matlab 2014	IEEE	2016
5	Storage Path Encryption Techniques, Homomorphic Encryption, Privacy Preserving Apriori Algorithm.	Splunk, Pentaho, Tableau, Talend	Springer	2016
6	Clustering method for Big Data,Visual analytics,Machine learning	Hadoop,SQL	IEEE	2016
7	HACE Theorem, 3DES Algorithm, Bayesian Algorithm.	Mahout, R-Hadoop	IEEE	2016
8	Agile Modelling Technique, User Context Modelling.	HHC System Platform, MOH Hajj App, Magnetic Card, Mobile Phones	IEEE	2016
9	K- means Algorithm, MaRCO Technique, Dynamic Decomposition	HDFS, Hadoop	IEEE	2016

	Algorithm.			
10	Map-Reduce Techniques, Big data analytics, Storage and management of big data Framework	Pig, Hive,HBase,HDFS,Hadoop	IEEE	2014

III. PROPOSED ARCHITECTURE

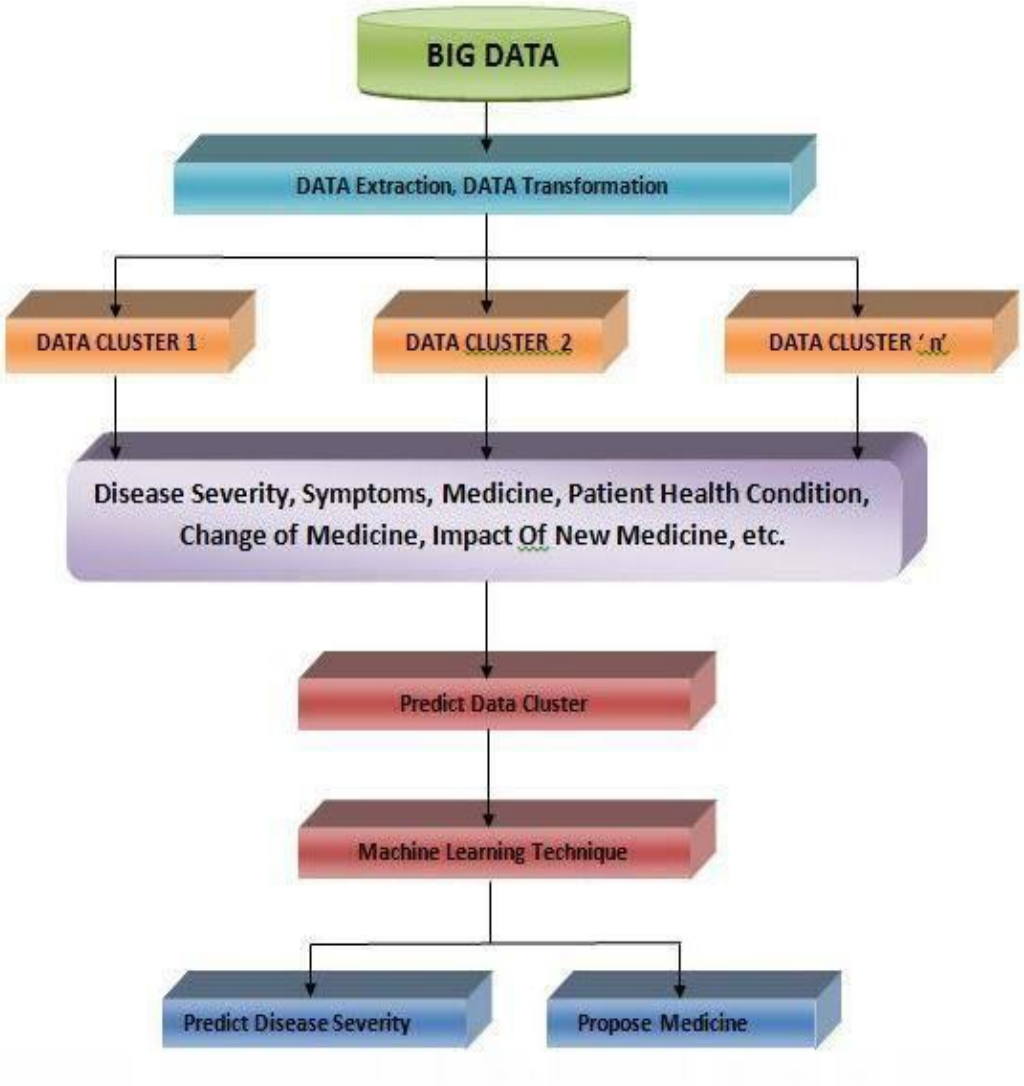


Fig 2 : Proposed Architecture

IV. EXPLANATION ON PROPOSED SYSTEM

The proposed system in this paper enhances the intelligent decision making phenomenon. The bulk amount of big data is extracted from database and transformed over data structure. The data then differentiated and forms cluster of respective sets. This clusters conducts all patients healthcare database. According to disease severity, symptoms, medicine, patient health condition, change in medicine for particular disease and impact of new medicine on this basis data get classified. In next stage proposed model prediction of data cluster and simultaneously necessary machine learning techniques are applicable. Hence at the end we are able to find disease severity and proposed medicine to cure it[5].

V. CONCLUSION

A. Health care related data care is huge in nature and they arrive from different places.

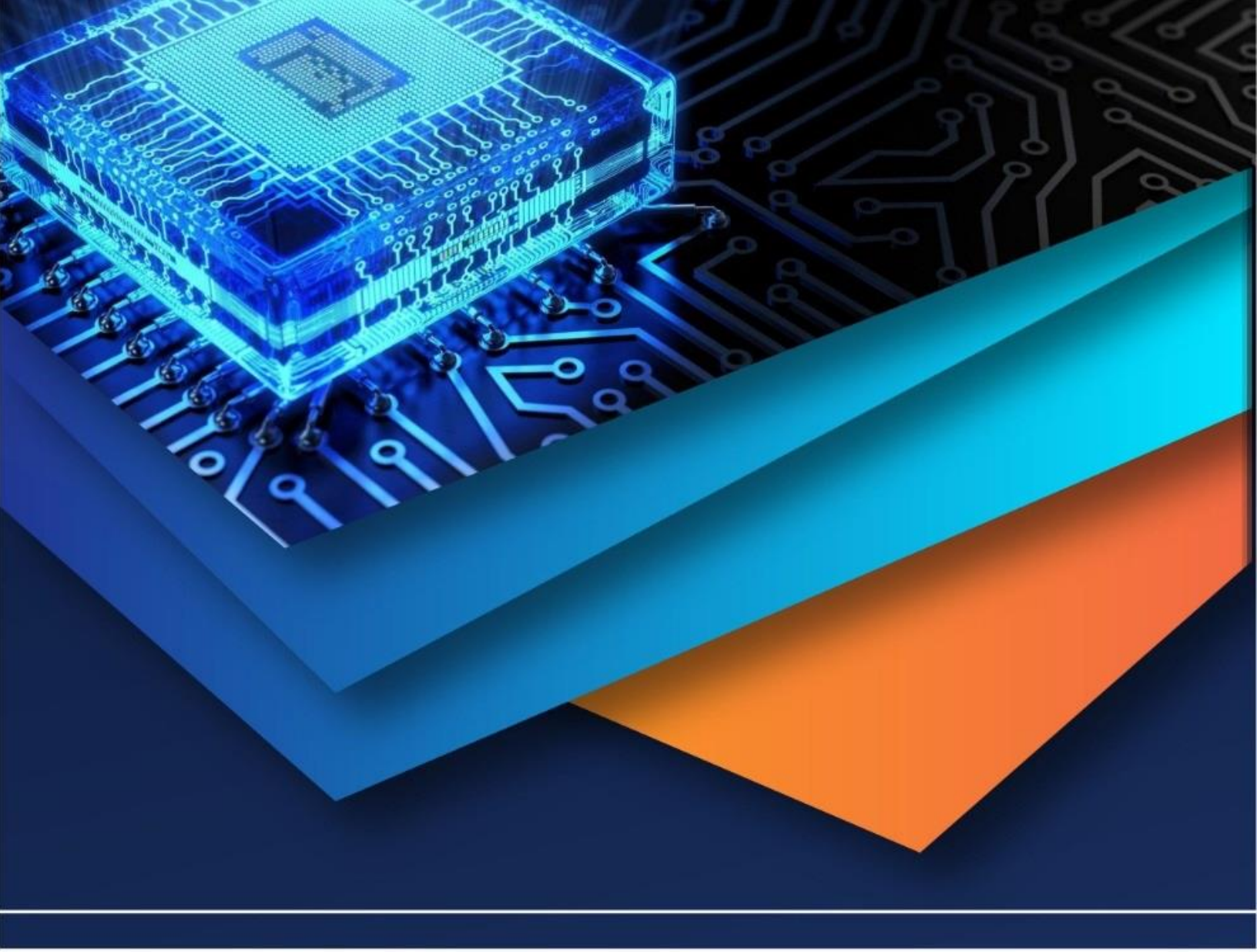
B. The utilization of knowledge and experience of specialists and medical screening data of patients are collected in a database during the diagnosis process, which has been widely accepted.

VI. ACKNOELEDGMENT

Special thanks to our Staff of Department of Computer Engineering, Jayawantrao Sawant College of Engineering & Sciences Hadapsar Pune and SavitribaiPhule Pune University in order to support this work.

REFERENCES

[1] R. Bhardwaj, A. Sethi, and R. Nambiar, "Big data in genomics: An overview," in Big Data (Big Data), 20i4 iEEE international Conference on, 2017, pp. 45-49.
[2] D. L. Rubin, S. E. Lewis, C. J. Mungall, S. Misra, M. Wester field, M.Ashburner, et al. , "National center for biomedical ontology: advancing biomedicine through structured organization of scientific knowledge," Omics: ajournal of integrative biology, vol. 10, pp. 185-198,2006
[3] T. W. Kim, K. H. Park, S. H. Yi, and H. C. Kim, "A Big DataFramework for u-Healthcare Systems Utilizing Vital Signs," in Computer, Consumer and Control (IS3C), 2014 International Symposium on, 2014, pp. 494-497.
[4] W. Raghupathi, "Data mining in health care," Healthcare Informatics: Improving Efficiencyand Productivity, pp. 211-223,2 0 I O
[5] B. Huppertz and A. Holzinger, "Biobanks-A source of large biological data sets: Open problems and future challenges," in Interactive Knowledge DiscovelY and Data Mining in Biomedicallriformatics, ed: Springer, 2016, pp. 317-330
[6] Muhammad Babar,MuhammadJehanzeb,F.Sher,D.jawawi, "Big Data survey in healthcare and a proposal for intelligent data diagnosis framework," UniversitiTesknology Malaysia International Islamic University of Islamabad Pakistan , 2016
[7] V. Kumar, H. Park, R. C. Basole, M. Braunstein, M. Kahng, D. H. Chau, et al., "Exploring Clinical Care Processes Using Visual and Data Analytics: Challenges and Opportunities," in Knowledge Discoveryand Data Mining (KDD): Workshop on Data Science for Social Good, 2016.
[8] A. Castiglione, R. Pizzolante, A. De Santis, B. Carpentieri, A. Castiglione, and F. Palmieri, "Cloud-based adaptive compression and secure management services for 3D healthcare data," Future Generation Computer Systems, vol. 43, pp. 120-134,2016.
[9] T. Song, "Efficient utilization of big data on healthcare and welfare area," in Health WeifPolicy Forum, 2016, pp. 68-76.
[10] U. Srinivasan and B. Arunasalam, "Leveraging Big Data Analytics to Reduce Healthcare Costs," 2016
[11] Mrs.VibhavariChavan,Prof.R.Phursule, "Survey Paper on Big Data,"2014**”**.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY