Available online at: http://www.ijmtst.com/vol5issue06.html

International Journal for Modern Trends in Science and Technology

ISSN: 2455-3778 :: Volume: 05, Issue No: 06, June 2019

Processing and Analyzing Healthcare Big Data on Cloud Computing Environment by Implementing Hadoop Cluster

Patel Vishruti K[1] | Prof. Madhav D. Ingle[2]

[1,2]Department of Computer Science, Jaywantrao Sawant College of Engineering, Pune, India.

To Cite this Article

Patel Vishruti K and Prof. Madhav D. Ingle, "Processing and Analyzing Healthcare Big Data on Cloud Computing Environment by Implementing Hadoop Cluster", *International Journal for Modern Trends in Science and Technology*, Vol. 05, Issue 06, June 2019, pp.-19-26.

Article Info

Received on 07-May-2019, Revised on 27-May-2019, Accepted on 10-June-2019.

ABSTRACT

In today's world, due to technological advancements, the amount of data that is getting generated is growing rapidly. Enterprises worldwide will need to perform data analytics with these huge data datasets to make business decisions and stay competitive. Storage of data sets and performing data analytics was traditionally accomplished using RDBMS (Relational Database Management System). However, RDBMS would be inefficient and time consuming when performing data analytics on huge data sets. A cloud based big data analytic platform is the best way to analyze the structured and unstructured data generated from healthcare management systems. Hadoop came into existence recently and overcomes the limitations of existing RDBMS by providing simplified tools for efficient data storage and faster processing times for data analytics. The purpose of this work is to perform data analytics on a health care data set using Hadoop functionalities. A health care data set comprising of 1.5 million patient records is considered for the data analysis. Different use cases as to be considered and will perform analytics using MapReduce, Hive and Pig functionalities of Hadoop.

KEYWORDS: *Big data, Hadoop, cloud, healthcare, prediction.*

I. INTRODUCTION

Data in health care industry is originated from different sources like clinical data, diagnosis data, doctor's prescription, electronic health records(HER), medical images, etc. collectively all this data becomes big data in health care industry. Big data is nothing but the data big in size. Healthcare insight the worldwide health care data grew up to 500 petabytes in 2012 and has been estimated to grow more than 25,000 petabytes by 2020. Big data applies to information which cannot be processed or analyzed using traditional tool and technique. To figure out value of the big healthcare data, to deliver the best evidence based, patient-centric, result oriented and accountable care. As suggested by AngappaGunasekaran[1] main characteristics of big data are volume, velocity, veracity, verity, Value. Volume indicates the amount of data that may in Gbs or Tbs or more than that. Velocity is consist of how fast the data are created and collected. Verity means how many types of data are present that may be structured, semi-structured or unstructured. Veracity means

the truthfulness of data. Value means, how much the data is valuable. Educating the superiority of health care and decreasing the cost is a principle behind the developing movement toward value based healthcare delivery model and patient-centered care. The volume and demand for big data in healthcare organizations are growing little by little [3]. The Big data processing technologies includes machine learning percentages determine the algorithms, natural language processing algorithms, predictive modelling and other artificial based techniques. Analysing Big Data will bring out better outputs for the organizations and will improve their performance. It can be seen from Figure. 2 that the term 'big data in healthcare' really took off around early 2013. Increase in interest in this term can be related to a

Popular report by McKinsey & Company that came out in early 2013 [10]. The report highlights that healthcare expenses contributes about17.6% of GDP and have a potential to reduce healthcare Spending by $300 billion to $450 billion. The figure 1 shows the different data to be diagnostic.

Fig1. Different Data Diagnostic

Important application of Big Data in the healthcare domain includes the Medical Body Area Networks (MBANs). MBANs allows a continuous monitoring of patient's condition by sensing and transmitting recorded measurements such as heart rate, electrocardiogram (ECG), body temperature, respiratory rate, chest sounds, and blood pressure etc. MBANs will allow:

1. Real-time and historical monitoring of patient's health;
2. Control of Infection;
3. Tracking and identification of patients.

II. LITERATURE SURVEY

It is found that developing an HMS that can alert the pharmacist of the expiry date of drugs at a given time and handle all departments in the hospital. Min Chen suggested that key benefit of the method came from: 1) The global interdependence model of EL decisions; 2) The purely collective nature of the MAP REDUCE algorithm [6]. Report submitted by U. S. Congress in August 2012 explains big data as "large volumes of high velocity, complex, and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information" [17]. The drawback for existing system is that does not provide an index mechanism. And does not support for multi structure data.

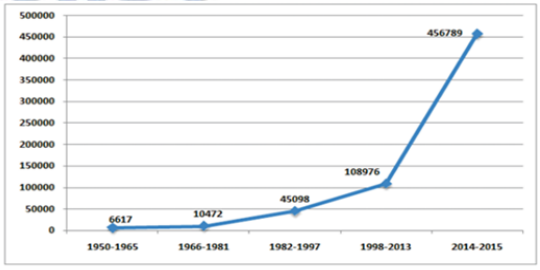


Fig. 2. Aggregate Number of Hospitals from 1950–2015.

Big data processing can be performed through two manners:batchprocessing and stream processing; see [4].Thefirstmethod is based on analyzing data over a specified period oftime; it is adopted when there are no constraints regardingthe response time. On the other hand, stream processing issuitable for applications requiring real-time feedback. Batchprocessing aims to process a high volume of data by collectingand storing batches to be analyzed in order to generateresults.Batch mode requires ingesting all data before processing it in a specified time.Mapreduce represents a widely adoptedsolution in the field of batch computing; see [5]; it operatesby splitting data into small pieces that are distributedto multiple nodes in order to obtain intermediate results.Once data processing by nodes is terminated, outcomeswill be aggregated in order to generate the final results.

Seeking to optimize computational resources use,mapreduceallocates processing tasks to nodes close to data location.Thismodel has encountered a lot of success in many applications, especially in the field of bioinformatics and healthcare. Batchprocessing framework has many characteristics such as theability to access all data and to perform many complex computationoperations, and its latency ismeasured byminutes ormore.Big data stream mining methods including classification,frequent pattern mining, and clustering relieve computationaleffort through rapid extraction of the most relevantinformation; this objective is often achieved by mining data inadistributedmanner. Thosemethods belong toone of the twofollowing

classes: data-based techniques and task-based techniques; [7]. Data-based techniques allow summarizing the entire dataset or selecting a subset of the continuous flowof streaming data to be processed. Sampling is one of these techniques; it consists of choosing a small subset of data to beprocessed according to a statistical criterion. Another data based method is load shedding which drops a part from theentire data, while sketching technique establishes a randomprojection on a feature set. Synopsis data structures methodand aggregation method belong also to the family of databased techniques.

Table 1 Literature survey

Sr No.	Title Of Project	Details of Publication	Algorithm	Pros	Cons	Example	Description
1.	Automated Health Care Management System Using Big Data Technology.	M. Ashish Reddy, Journal of Network Communications and Emerging Technologies, April (2016).	Map Reduce Algorithm	Solved problem of data quality in electronic patient records using a compute-rized patient records report system.	Semi-structured data.	Doctor's module and Pharmacist's module	Developing an HMS that can alert the pharmacist of the expiry date of drugs at a given time and handle all departments in the hospital.
2.	Smart Clothing: Connecting Human with Clouds and Big Data for Sustainable Health Monitoring.	Min Chen. Springer Science And Business Media New York 2016.	Novel participant selection algorithm.	Good accuracy	slow training, computationally expensive	children's health, image-based MR classification	Introduces a system which can promote active and healthy lifestyles, and provide people with valuable healthy lifestyle and habits.
3.	Big Data for Health	Javier Andreu-Perez, Carmen C. Y. Poon, Robert D. Merrifield, Stephen T. C. Wong, IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, JULY 2015	Biostatistics and Genomics.	Increased data processing power.	Single sensing element.	Gene expression data.	Study of gene networks of different syndromes of the same person in Order to better understand how these syndromes are interrelated.
4.	Medicine and physical health system Using IoT.	P.Elanthiraiyan International journal march 2015.	Density-based Clustering	handle non-static and complex data, detect outliers and arbitrary shapes	slow, tricky parameter selection, not well for large datasets	biomedical image clustering, finding bicliques in a network	The Author studied the resources optimizations and how to monitor the data using IOT related to health care
5.	Applications of the MapReduce programming framework to clinical big data analysis: current landscape and future trends.	Emad A Mohammed , Behrouz H Far, Biodata Mining 2014	(SNP) selection algorithms and Genome sequence comparison algorithm.	reliable data processing.	Does not provide an index mechanism.	Clinical datasets.	Demonstrates the significance of using the MapReduce programming model on top of the Hadoop distributed processing platform to process the large volume of clinical data.
6.	Telemedicine-Based WBAN Framework for Patient Monitoring.	Chinmay Chakraborty MARY ANN LIEBERT, INC. AUGUST 2013 TELEMEDICINE and e-HEALTH	Low-energyAdaptiveClustering Hierarchy(LEACH),Power-efficient Gathering in SensorInformation Systems (PEGASIS.	Provide patient health data in real-time.	link loss in the network in the Healthcare domain.	heart problems, emergency response, asthma, deep brain stimulation.	The main function of this unit is to collect patient physiological data and Forward them to the medical centre in an efficient and reliable way.
7.	Research and Implementation of Massive Health Care Data Management and Analysis Based on Hadoop.	Hongyong Yu, Deshuai Wang 2012 Fourth International Conference on Computational and Information Sciences.	Traditional data analysis methods.	better perform-ance, scalability	data security	massive health care data.	Result prove that the Hadoop based framework improves the performance of data upload and data query dramatically, and the Hive-based data analysis method is suitable for massive data analysis tasks.
8.	Framework for Dynamic Integration of Multimedia Medical Data Into Distributed m-Health Systems.	Liviu Constantinescu, IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, VOL. 16, NO. 1, JANUARY 2012.	Centrality Measure-ment and SparkMed Framework	Pervasive Automated network for Medical Enterprise Data.	Limited scope of access to data in proprietary hospital infrastructure systems.	Hospital imaging software.	Health, is one of the fastest growing areas of healthcare computing As electronic health records become Common place, and the rapid uptake of Mobile and handheld devices puts powerful.
9.	Implementation of a Medical Image File Accessing System on Cloud Computing.	Chao-Tung Yang Lung-Teng Chen, 2010 13th IEEE International Conference on Computational Science and Engineering.	HDFS file system.	Easy Manage-ment and Cost effective.	Single point of failure.	Medical image files.	presented a system called MIFAS (Medical Image File Accessing System) to solve the exchanging, storing and sharing on Medical Images of crossing the different hospitals issues.
10.	Big Data Solutions in Healthcare: Problems and Perspectives	Prabha Susy Mathew,, Dr.Anitha S. Pillai, IEEE Sponsored 2nd International Conference on Innovations in Information Embedded and Communication Systems.2015	data mining algorithms, SAP HANA with in-memory computing	data can be highly compressed and it is self indexing and uses less space compared to RDBMS.	time issue.	Clinical Data	With convergence of advanced computing and numerous Big Data technological options like commercial solutions, Open Source, Cloud etc. it is now possible to attain high performance, scalability at a relatively low cost.

III. PROPOSED METHODOLOGY

The proposed system as a Hadoop framework provides four core modules:

1. The inbuilt functions are used by the Hadoop modules.
2. HDFS: Hadoop Distributed File System, It is a java based file system, the data can be expanded and can store data across different machines in an organization.
3. It provides the software model for processing large data sets in parallel.
4. It provides YARN (Yet another Resource Negotiator) framework, for managing, scheduling and handling the resources from middleware applications [12].Big data provides a great opportunity for epidemiologists, physicians, and health policy experts to make data-driven judgments that will eventually develops the patient care [8].The author D. I. Sessler have used Google trends for analyzing the 'big data in healthcare' between 2010 and 2015. The resulting graph is shown in figure 2.

A. Architecture

Data are stored in the HDFS and made available to the slave nodes for computation. Author W. Zhang mention that the Prediction begins with the identification of symptoms and dataset classified for that in a way to detect the disease[6]. The Figure 4 presents the proposed system for data analysis in healthcare management.

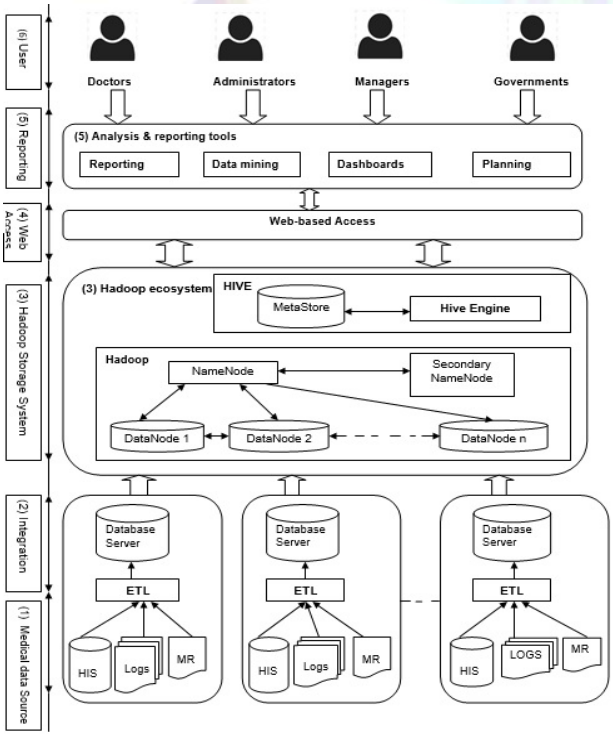


Fig 3. Hadoop-based system architecture of medical big data

The proposed tool enables agencies too easily and economically clean, characterize and analyze the data to identify actionable patterns and trends. There are various Big Data handling tools medical data but we choose Hadoop, since it is an open source and it can handle huge amount of unstructured data. Interesting direction will be to build patient-specific models using data already available in existing clinical databases, and, then update the model with data that can be collected outside the hospitals. Health informatics applications are known to generate datasets that are complicated to store, untangle, organize, process, and, above all, interpret. The data mining process for identifying the most effective mode of treatment for each age group, particularly for younger and older age patients, was divided into six steps. The processing blocks are shown in Fig. 3:'Data Mining Architecture of Disease'.

A. Data selection: The first stage of the mining process is data selection. In this step, the data are prepared and errors such as missing values, data inconsistencies, and wrong information are corrected.

B. Data preparation: The data preparation stage is crucial for data analysis. Databases stored in .xls or .mdb format were found to be insufficient. The Oracle Data Miner software re-quires input to be provided in a particular format. Consequently, it was deemed necessary to convert the database to Oracle Database 10g format to facilitate use with the Oracle Data Miner.

C. Data analysis: In the data analysis stage, data are analyzed to achieve the desired research objectives, for example by selecting the appropriate target values from the master table. In a data mining engine, the data mining techniques comprise a suite of algorithms such as SVM, Naive Bayesian, etc. In this study, we used a regression technique that employed a support vector machine algorithm.

D. Result database: At this stage, the desired algorithm and associated parameters have been chosen. The Oracle Data Miner software has a specific option, such 'publish', and that processes the raw data and creates a result database.

E. Knowledge evaluation and pattern prediction: This stage extracts new knowledge or patterns from the result database. An informative knowledge database is generated that facilitates pattern forecasting on the basis of prediction, probabilities, and visualization.

F. Deployment: The final stage of this process applies a previously selected model to new data to

generate predictions

B. Algorithm

1) The Inter-Cluster Correlation Evaluation has been implemented by P. K. Sahoo et al[12] In which author mention the Inter-cluster correlation algorithm that allows to find the similarity or dissimilarity between health parameters of different departments. With the inter cluster correlation algorithm to analysis the *J48 and SVM* algorithm we can use with advancement of Apache Hive. The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space (N—the number of features) that distinctly classifies the data points health nodes.The support vector machine is a training algorithm for learning classification and regression rules from data. The SVM is based on statistical learning theory. The SVM solves the problem of interest indirectly, without solving the more difficult problem. The support vector machine presents a partial solution to the bias variance trade-off dilemma. There are two ways of implementing SVM. The first technique involves mathematical programming and the second technique employs kernel functions. When kernel functions are used, SVM focuses on dividing the data into two classes, P and N, corresponding to the case when yi= +1 and yi=1, respectively. The support vector classification searches for an optimal separating surface, called a hyperplane, which is equidistant from each of the classes. This hyperplane has many important statistical properties and kernel functions are non-linear decision surfaces (Burbidge and Buxton, 2001). If training data are linearly separable, then a pair (w,b) exists such that,

$$\mathbf{w}^T\mathbf{x}_i + b \geq 1 \text{ for all } \mathbf{x}_i \in P, \text{ and}$$

$$\mathbf{w}^T\mathbf{x}_i + b \leq -1 \text{ for all } \mathbf{x}_i \in N$$

Where w is a weight vector and b is a bias. The prediction rule is given by:

$$f = \text{sign}(\langle \mathbf{w} \cdot \mathbf{x} \rangle + b).$$

2) Hadoop Partitioner

The Partitioner in MapReduce controls the partitioning of the key of the intermediate mapper health data output. By hash function, key (or a subset of the key) is used to derive the partition among all the available data node. MapReduce job takes an input data set and produces the list of the key-value pair which is the result of map phase in which input data is split and each task processes the split and each map, output the list of key-value pairs[18]. Then, the output from the map phase is sent to reduce task which processes the user-defined reduce function on map outputs. But before reduce phase, partitioning of the map output take place on the basis of the key and sorted. This partitioning specifies that all the values for each key are grouped together and make sure that all the values of a single key go to the same reducer, thus allows even distribution of the map output over the reducer. Partitioner in Hadoop MapReduce redirects the mapper health record output to the reducer by determining which reducer is responsible for the particular key.

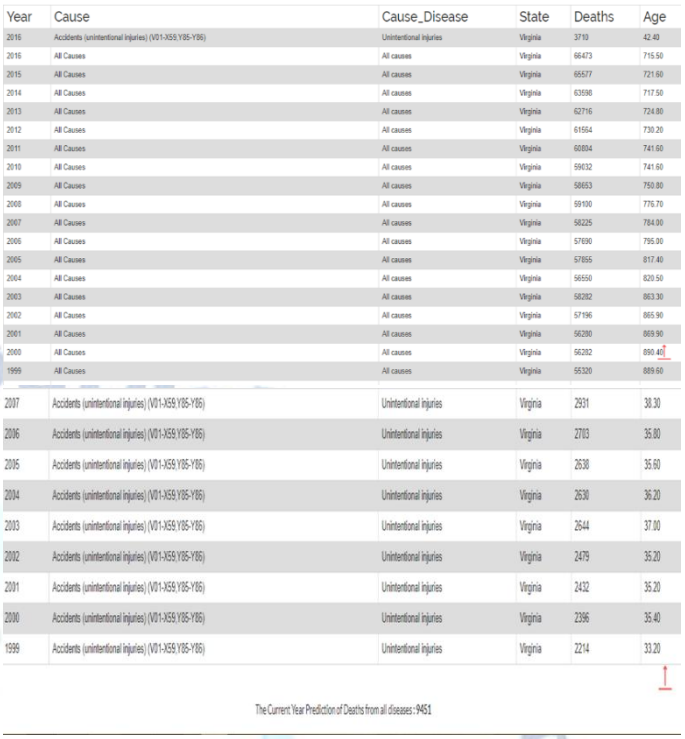
IV. EXPRIMENTAL RESULT AND DISCUSSIONS

The SHIH-LIN WU implemented the algorithms [12] which carried out by using CloudSim for the health data predication. Number of heart disease attributes are used which as the structured data. The result for the data was around the 51%. The current implementation for our work include the prediction for the structure, unstructured and antistructure data.The five age groups were re-classified into two age groups: Young and Old. Predictions based on the young group and old age groups were denoted as p(y) and p(o), respectively. The p(y) group included the 15–24, 25–34 and 35–44 age groups, while the p(o) group included the 35–44, 45–54 and 55–64 age groups. It should be noted that the '35–44' group is common to both of the two age groups. This implies that drug treatment is more effective for patients in the old age group than patients in the young group. Apart from medication, exercise, weight reduction and smoke cessation are also important aspects of effective treatment. In our simulation, the input data sizes are set in gigabytes that range from 5GB to 50GB. In a current section of implementation include the cluster which introduce of how to use the framework to address the problem of medical resources distribution. Indeed, we used the set of standard data set NCHS_United_State.arff National Center for Health Statistics Repository with records and attributes.. The standard data set used to analyase the record is Electronic Health Record Incentive Program Providers. However, using time series data, we predict the number of expected patients in the next twelve months by using the historical data of 19 years i.e. April 1999-18 to February 2017-19 with AWS Modeler. In addition to this, the lower interval and upper

interval range of data is also predicted for the next twelve months from September 2018 to Oct 2019. Similarly, we can predict the future data for 2-5 years based on the historical data. Time series data mining is an integrated solution to forecast correct results that are totally based upon the accurate historical data. So time series data mining offers assurance in helping organizations to uncover hidden patterns in their data. The effectiveness of the proposal is evaluated by conductingexperiments with a cluster formed by 3 nodes withidentical setting, configured with an Intel COREi7-4770processor (3.40GHZ, 4 Cores, 16GB RAM, running Ubuntu12.04 LTS with 64-bit Linux 3.11.0 kernel) as shown is figure 9.Medical care of such patients is achallenging process since a lot of checks are performedmany times during a single day; for instance, some diabeticsmeasure their blood pressure several times on a daily basis.

Fig4. Training and testing data

It give two reports; the first one is based on 'Hospitalization' fact table, and the second one is based on 'Consultation' fact table. Figure 5 shows one of the first results of the reporting phase which consists of a comparison between of three data node with in the university hospital and the five public hospital institutions of different cities available in the state of California. The proposed system include certain advantage such as 1) Reduce the large amount time for generating report. 2) Effectively handle voluminous and heterogeneous healthcare data. 3) Detect the most common diseases among patient. Some application which help our society to provide quantitative information in variety of applications such as disease stratification, predictive modeling, and decision making systems.



Year	Cause	Cause_Disease	State	Deaths	Age
2016	Accidents (unintentional injuries) (V01-X59,Y85-Y86)	Unintentional injuries	Virginia	3710	42.40
2016	All Causes	All causes	Virginia	66473	715.50
2015	All Causes	All causes	Virginia	65577	721.60
2014	All Causes	All causes	Virginia	63698	717.50
2013	All Causes	All causes	Virginia	62716	724.80
2012	All Causes	All causes	Virginia	61564	730.20
2011	All Causes	All causes	Virginia	60804	741.60
2010	All Causes	All causes	Virginia	59632	741.60
2009	All Causes	All causes	Virginia	58653	750.80
2008	All Causes	All causes	Virginia	59100	776.70
2007	All Causes	All causes	Virginia	58225	784.00
2006	All Causes	All causes	Virginia	57690	795.00
2005	All Causes	All causes	Virginia	57855	817.40
2004	All Causes	All causes	Virginia	56550	820.50
2003	All Causes	All causes	Virginia	56282	863.30
2002	All Causes	All causes	Virginia	57196	865.90
2001	All Causes	All causes	Virginia	56280	869.90
2000	All Causes	All causes	Virginia	56282	890.40
1999	All Causes	All causes	Virginia	55320	889.60
2017	Accidents (unintentional injuries) (V01-X59,Y85-Y86)	Unintentional injuries	Virginia	2931	38.30
2016	Accidents (unintentional injuries) (V01-X59,Y85-Y86)	Unintentional injuries	Virginia	2703	35.80
2015	Accidents (unintentional injuries) (V01-X59,Y85-Y86)	Unintentional injuries	Virginia	2638	35.60
2014	Accidents (unintentional injuries) (V01-X59,Y85-Y86)	Unintentional injuries	Virginia	2638	36.20
2013	Accidents (unintentional injuries) (V01-X59,Y85-Y86)	Unintentional injuries	Virginia	2644	37.00
2012	Accidents (unintentional injuries) (V01-X59,Y85-Y86)	Unintentional injuries	Virginia	2479	35.20
2011	Accidents (unintentional injuries) (V01-X59,Y85-Y86)	Unintentional injuries	Virginia	2432	35.20
2010	Accidents (unintentional injuries) (V01-X59,Y85-Y86)	Unintentional injuries	Virginia	2396	35.40
1999	Accidents (unintentional injuries) (V01-X59,Y85-Y86)	Unintentional injuries	Virginia	2214	33.20

The Current Year Prediction of Deaths from all diseases : 9451

Fig5. Classification of diseases for virgina state

Total 40 states can be classified by using J48 algorithm and SVM classifier. Given above figure 5 shows the predictive analysis for the virgina state of USA. Correctly classified instances and incorrectly classified instance for the given data is analysed using predictive methodology i.e SVM. We get 45 total number of instances from ten thousand five hundred record. For those 45 instance firstly tree is generated using J48 Pruned tree. Later Decision list and decision table are formed.

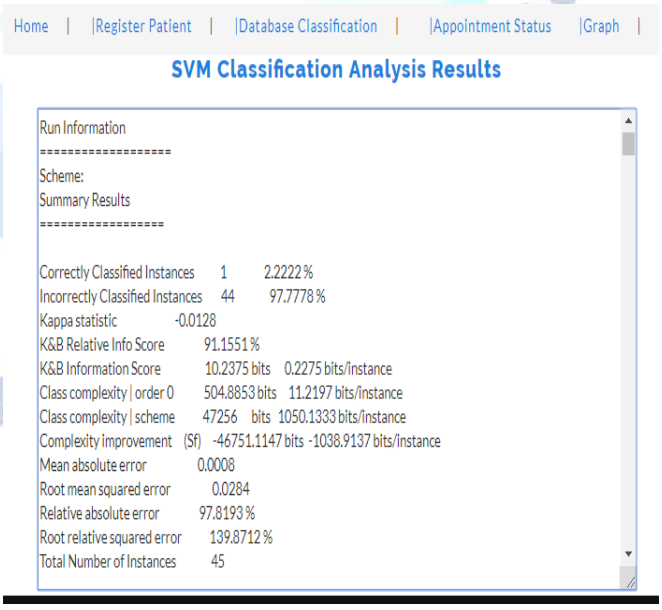


Fig6. SVM classification analysis

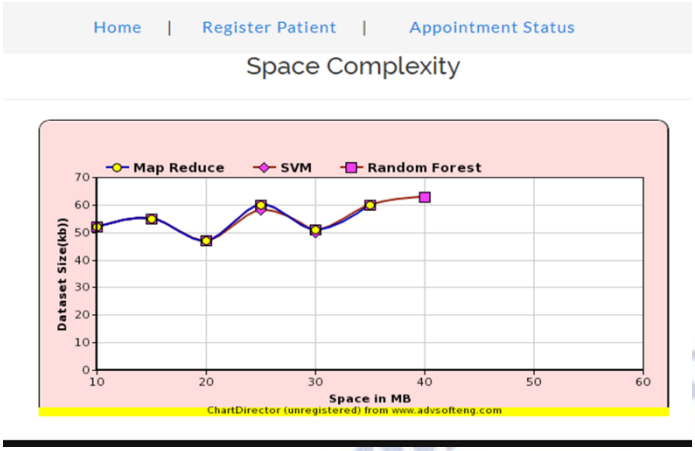


Fig7. Space Complexity

Average Analysis of Algorithm

Algorithm	Cluster Formation(ms)	Training/Testing(ms)	Accuracy(%)
MAP REDUCE	12000	--	70%
SVM	9700	10000	81%
Random Forest	140000	18000	75%

Fig8. Comparative Result

Data Management

Data Management in healthcare includes organizing, cleaning, retrieval, data mining, and data governance. It also includes the method of validating whether there is some scrap data or any missing values. Such data needs to be removed. It helps in risk assessment of patients, personalized discharge plan. Major data management tools are Apache Ambari and HCatalog. Data retrieval is a process of extracting file or valuable information from large healthcare databases. AWS cloud is used to monitor entire data.

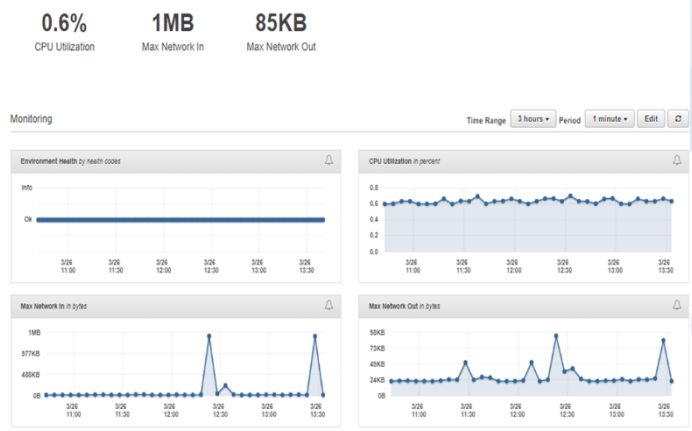


Fig9. System Monitoring Report on AWS Cloud

Big data analysis in healthcare frequently contains information recovery and data mining "information retrieval" is the process of searching within large document collections, and in healthcare it mainly covers medical text retrieval and medical image retrieval. Data governance refers to overall management of security, integrity, usability and availability of the data employed in an enterprise. Maintaining confidentiality of individual patient records is very important in Healthcare management.

V. CONCLUSION AND FUTURE WORK

Health care analysis is important to find and there's no better way to collect direct feedback regarding the patients and improve the product or service. However, the way of handling a patient's health using the healthcare analysis is simpler task to provide a finer health quality to the patients as earlier loading large amount of data is very difficult. By using Big data complexity of loading large amount of data can be reduced. The proposed tool enables agencies too easily and economically clean, characterize and analyze the data to identify actionable patterns and trends. There are various Big Data handling tools which can provide us information from unstructured data like medical data but we choose Hadoop, since it is an open source and it can handle huge amount of unstructured data. The information helps us in improving the different areas within a particular region. The information can also be helpful in designing and implementation of healthcare systems in a particular country. The relative time is also reduced and measured as 1x for 1 TB and 10 TB of data. Future research can include First, data security is very important in all data management systems. On large data processing, operations are often running remotely in the server or cloud side, new encryption method is needed to ensure the data privacy. And real time data analysis can be continued in future.

REFERENCES

[1] Big Data for Health AngappaGunasekaran,Javier Andreu-Perez, Carmen C. Y. Poon, Robert D. Merrifield, Stephen T. C. Wong, IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, JULY 2015

[2] Framework for Dynamic Integration of Multimedia Medical Data into Distributed m-Health Systems. Liviu Constantinescu, IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, VOL. 16, NO. 1, JANUARY 2012.

[3] M. Larsen, T. Boonstra, P. Batterham, B. O'Dea, C. Paris, and H. Christensen, "We feel: Mapping emotion on Twitter," IEEE J. Biomed. Health Informat., pp. 1–7, 2015.

[4] R. Nambiar, R. Bhardwaj, A. Sethi, R. Vargheese, A look at challenges and opportunities of Big Data analytics in

healthcare, Proc. - 2013 IEEE Int. Conf. Big Data, (2013) 17–22. doi:10.1109/BigData.2013.6691753.

[5] Implementation of a Medical Image File Accessing System on Cloud Computing. Chao-Tung Yang Lung-Teng Chen, 2010 13th IEEE International Conference on Computational Science and Engineering.

[6] S. Ram, W. Zhang, M. Williams, and Y. Pengetnze, "Predicting asthmarelated emergency department visits using big data," IEEE J. Biomed. Health Informat., pp. 1–8, 2015, submitted for publication.

[7] Changqing Ji, Yu Li, WenmingQiu, UchechukwuAwada, Keqiu Li. 2012 IEEE International Symposium on Pervasive Systems, Algorithms and Networks, PP 17-23, 2012.

[8] Big Data Solutions in Healthcare: Problems and Perspectives. D. I. Sessler, Prabha Susy Mathew, Dr.Anitha S. Pillai, IEEE Sponsored 2nd International Conference on Innovations in Information Embedded and Communication Systems.2015.

[9] Jing Bi, Zhiliang Zhu, Ruixiong Tian, and Qingbo Wang. 2010 IEEE 3rd International Conference on Cloud Computing (CLOUD), Miami, Florida, PP 370-377, 2010.

[10] M. Viceconti, P. Hunter, R. Hose, Big Data, Big Knowledge: Big Data for Personalized Healthcare, IEEE J. Biomed. Heal. Informatics. 19 (2015) 1209–1215. doi:10.1109/JBHI.2015.2406883.

[11] L. Wang, R. Ranjan, J. Kołodziej, A. Zomaya, L. Alem, Software Tools and Techniques for Big Data Computing in Healthcare Clouds, IEEE. (2015) 38– 39. doi:10.1016/j. future.2014.11.001.

[12] SHIH-LIN WU, P. K. Sahoo et al.: Analyzing Healthcare Big Data With Prediction for Future Health Condition, IEEE 2017.

[13] S. Kaisler, F. Armour, J. A. Espinosa, W. Money, Big Data: Issues and Challenges Moving Forward, IEEE Comput. Soc. 46th Hawaii Int. Conf. Syst. Sci. (2013) 995–1004. doi:10.1109/HICSS.2013.645.

[14] O. S. Lupse, M. Crisan-Vida, L. Stoicu-Tivadar, E. Bernard, Supporting diagnosis and treatment in medical care based on big data processing, Studies in Health Technology and Informatics. 197 (2014) 65–69.

[15] M. L. Berger and V. Doban, Big data, advanced analytics and the future of comparative effectiveness research, Journal of Comparative Effectiveness Research. 3 (2014) 167–176.

[16] B. Feldman, E. M. Martin, T. Skotnes, Big Data in Healthcare- Hype and Hope, Dr. Bonnie 360 degree (Business Development for Digital Health), 2012. http://www.riss.kr/link?id=A99883549.

[17] W. Raghupathi, V. Raghupathi, Big data analytics in healthcare: promise and potential, Heal. Inf. Sci. Syst. 2 (2014) 1–10. Doi:10.1186/2047-2501-2-3.

[18] Applications of the MapReduce programming framework to clinical big data analysis: current landscape and future trends. Emad A Mohammed, Behrouz H Far, Biodata Mining 2014

[19] Smart Clothing: Connecting Human with Clouds and Big Data for Sustainable Health Monitoring. Min Chen. *Springer Science and Business Media New York 2016.*

[20] Automated Health Care Management System Using Big Data TechnologyM. Ashish Reddy, Journal of Network Communications and Emerging Technologies, April (2016).

[21] A. G. Erdman, D. F. Keefe, Grand challenge: Applying regulatory science and big data to improve medical device innovation, IEEE Biomed. Eng. 60 (2013) 700-706. doi: 10.1109/TBME.2013.2244600.

[22] R. S. Kovats and S. Hajat, "Heat stress and public health: A critical review," in Annual Review of Public Health, vol. 29. Palo Alto, CA, USA: Annual Reviews, 2008, pp. 41–57.

[23] Big Data for Health Javier Andreu-Perez, Carmen C. Y. Poon, Robert D. Merrifield, Stephen T. C. Wong, and Guang-Zhong Yang, IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. 19, NO. 4, JULY 2015

[24] R. J. Hijmans, S. E. Cameron, J. L. Parra, P.G. Jones, andA. Jarvis, "Veryhigh resolution interpolated climate surfaces for global land areas," Int.J. Climatol., vol. 25, pp. 1965–1978, Dec. 2005.

[25] M. A. Friedl, D. Sulla-Menashe, B. Tan, A. Schneider, N. Ramankutty,A. Sibley, and X. M. Huang, "MODIS Collection 5 global land cover:Algorithm refinements and characterization of new datasets," Remote Sens. Environ., vol. 114, pp. 168–182, Jan. 2010.

[26] S. Moltchanov et al., "On the feasibility of measuring urban air pollutionby wireless distributed sensor networks," Sci. Total Environ., vol. 502,pp. 537–547, 2015.

[27] B. Lobitz, L. Beck, A. Huq, B. Wood, G. Fuchs, A. S. G. Faruque, andR. Colwell, "Climate and infectious disease: Use of remote sensing for detection of Vibrio cholerae by indirect measurement," Proc. Nat. Acad. Sci. United States Amer., vol. 97, pp. 1438–1443, Feb. 2000.

[28] G. Luber and M. McGeehin, "Climate change and extreme heat events,"Amer. J. Prev.Med., vol. 35, pp. 429–435, Nov. 2008.

[29] J. C. Semenza and B. Menne, "Climate change and infectious diseasesin Europe," Lancet Infectious Dis., vol. 9, pp. 365–375, Jun. 2009.

[30] J. C. Eichstaedt, H. A. Schwartz, M. L. Kern, G. Park, D. R. Labarthe, R.M. Merchant, S. Jha, M. Agrawal, L. A. Dziurzynski, M. Sap, C. Weeg, E. E. Larson, L. H. Ungar, andM. E. Seligman, "Psychological language on Twitter predicts county-level heart disease mortality," Psychol. Sci., vol. 26, pp. 159–69, Feb. 2015.

[31] Raja, P.V., and Sivasankar, E., Modern Framework for Distributed Healthcare Data Analytics Based on Hadoop. In Information and Communication Technology-EurAsia Conference (pp. 348–355).

[32] Mackenbach, C. McDonald, S. Nayha, and I. Vuori, "Cold exposure and winter mortality from Ischaemic heart disease, cerebrovascular disease, respiratory disease, and all causes in warm and cold regions of Europe," Lancet, vol. 349, pp. 1341–1346, May 1997

[33] R. Nambiar, R. Bhardwaj, A. Sethi, R. Vargheese, A look at challenges and opportunities of Big Data analytics in healthcare, Proc. - 2013 IEEE Int. Conf. Big Data, (2013) 17–22.

[34] U. S Government, Department of Health and Human Services, Fedral Register, Rules and Regulations, 74(2009) 56123- 56131,

Availablefrom:https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/administrative/enforcementrule/enfir.