

An Efficient Hadoop Based Healthcare Data Mining Services

Dr.A.R.PonPeriasamy¹, Mrs.M.Chandamona ²

¹ Associate Professor, Nehru Memorial College,Puthanampatti,Trichy

² Assistant Professor & Head of the Department BCA, G.T.N Arts College,Dindigul

Abstract - The emergence of latest data technology like mobile application, cloud computing, big data Analytics compact all sectors. This can be significantly true for the health care system as a crucial sector, Nowadays Healthcare trade depends primarily on data technology to produce best services. Which promise the healthcare space a giant modification particularly ahead of the explosion of medical data sources: data analysis and also the extraction from medical data victimization data mining techniques. We will say that a big-data revolution is under approach in health care and begin with the immensely enlarged offer of health data. Which push us to apply these new technologies to induce off their benefits and improve the medical sector. This paper can show the importance of applying predictive analytics techniques in medical platforms, and provides architecture design which mixes big data analysis, data-mining and also the mobile health care for self-monitoring. this method are going to be ready to exploit the attention data through an intelligent method analysis and big data processing; so as to extract helpful information to serving to in decision making and guarantee a medical monitoring in real-time.

Key Words: a platform of healthcare data analysis, healthcare data, big data analytics; data mining, predictive analytics.

1.INTRODUCTION

Today digitized data is present everyplace as a result of information is growing and moving quicker than health care organizations will consume it. This can be due primarily to the efforts of researchers within the medical field and their discoveries take as an example human DNA. Widespread use of the electronic medical records would like completely transforms medical aid [7].

The newest innovations regarding genetic science and smart home or smart places allows patient self-monitoring and treatment by victimization less complicated devices [15]. the appearance of sensing technology like M-health [11]; health care data with all of this becomes voluminous and looks like a digital flood making puddles and lakes, creeks and torrents, of data: numbers, words, voices, images, video; and this will increase in parallel with the rapid growth within the use of mobile devices like smartphones, laptops, tablets, personal sensors that generating a data deluge.

Large data volumes at high velocities were originally a possibility that characterizes supercomputers, nuclear physics, military simulations and space travel. Late within the twentieth century, bigger and faster data appeared in airline and bank operations, significantly with the expansion of credit cards. Beginning in 1990, The Human ordination Project was the launch of

Big data in tending [21], and this was as a result of a statistic that showed that eightieth of medical information is unstructured and is clinically relevant and far important. This data resides in multiple places like individual EMRs, workplace and imaging systems, physician notes, medical correspondence, claims, CRM systems, and finance. For that a data-intensive analysis effort that pushed the bounds of obtainable data processing technology.

The potential of massive data analytics permits us to hope to slow the ever-increasing costs of care, facilitate providers to practice more practical drugs, empower patients and health care providers, support fitness and preventive self-care, and to dream regarding a lot of personalized and predictive medication. Yet, like the web, social media, cloud computing, and victimization the intelligent procedure for managing analyzing and extracting information from Data; all which will rework healthcare system and provides the ability to explore , predict and why not anticipate the cure. Big-data analysis guarantees and affirms that future isn't any longer mysterious.

As follows we'll discuss the good role contend by new technology within the field of health like health care analysis, and then we'll present our proposed design and its contribution in health care sector.

The rest of this paper can present as follow: in section II, we tend to present related works regarding technologies applied in the health care system and researches add this field. Section III is reserved for description of our proposed platform and the last section provides conclusions and views.

2.RELATED WORK

If we tend to mention Cloud computing as new technology applied within the health care system ,it brings several advantages ; by making a network of doctors; patients and health care institutes and facilitates access to medical data anyplace and anytime [1],

Cloud computing provides health care much-appreciated services regarding data handling by making certain [2, 3]:

- Resiliency: cloud service providers offer platforms with a awfully powerful infrastructure that has redundancy and storage of any data amount to making sure high accessibility anytime and anyplace.
- Privacy: cloud computing infrastructure guarantees a high level of security than local IT department during a hospital will ensure.
- Speed of innovation: everything is handled within the cloud data redundancy and also the update. By cloud provider don't want doing updates or putting in the certificates or repairing interference systems
- Mobile applications: whereas the mobile applications used are keep within the cloud and also the data is additionally hold on within the cloud; the communication will be exhausted a better and additional flexible approach as long as the ability of access are identical to at least one patient or many at a similar time .
- Developing trend: cloud adapts to any or all things to make sure easy access at a high level.

A lot of researchers are centered during this regard [14, 10] and it cited the big| role played by cloud computing within the stage of managing tending data that have become progressively large. More than, that a number of them provide design of a cloud computing-based health care SaaS Platform (HSP) to deliver health care data services with low cost, high clinical worth and high usability with a high level of security [8, 6]. big data analysis particularly in health care area has been thought of as a revolutionary approach to improving the standard of health care service [4, 9], as a result of analytics figures to play an important role within the way forward for health care system and as a results of analysis to develop tending sector [21, 22] systems found obliged to receive a new form of data like human DNA, data genetics; thus the requirement of investing of these resources and embitter human health. Analytics also can be applied in health care to check the value and effectiveness of interventions, treatments, public health policies, or medical devices to reduce failed investments. In fact, this sort of research might offer the simplest solution to prevent medical disasters. As an example, infectious diseases may be predicted by data health care analysis and also the health authority may manage this case and save the human. Can presently be overflowing in genomic data [5, 24]. Given the incredible size and spatial property of those datasets, the field of analytics can got to borrow techniques to form it helpful. Additionally to it, some predictive analytics platforms for disease targets across varied patient cohorts

victimization electronic health records (EHRs) are created to facilitate specific biomedical analysis workflows, like refinement of hypotheses or data semantics [13].

Regarding tools employed in that the foremost necessary platform for big data analytics platform is that the open-source distributed processing platform Hadoop (Apache platform) [20]. It belongs to the category of technologies "NoSQL" that have evolved to managing data at high volume. Hadoop has the potential to method extraordinarily large and primarily by allocating partitioned off data sets to varied servers (nodes), every of that solves completely different elements of the larger downside and so integrates them for the ultimate result [29-30]. Hadoop will serve each roles of organizing and data analyzing tool .Hadoop will handle terribly giant volumes of data with completely different structures or no structure in any respect. However Hadoop could be a very little difficult to install, configure and manage, and people with Hadoop skills don't seem to be simply found. Additionally, for these reasons, it seems organizations aren't quite able to embrace Hadoop completely. Knowing that the adoption of EHRs and electronics data, ready a submitted base for applying analysis and become the norm in health care, it permits the building of predictive analytic solutions.

These predictive models, as we all know have the potential to lower value and improve the general health of the population. As predictive models become additional pervasive, some standards seems to be utilized by all the parties concerned within the modeling process: just like the predictive Model markup language (PMML) [19].It permits for predictive solutions to be simply shared between applications and systems. And it are often accustomed expedite the adoption and use of predictive solutions within the health care industry.

Consistent with our analysis, we tend to found that there are several efforts to making platforms supported cloud computing for managing medical records and modify access to data. The patient doesn't care regarding the means together with his doctor manages his medical data. However that is additional vital to him is what's the positive impact of this on his health scenario on one hand, and on the opposite hand and get entangled within the treatment method. What we tend to propose could be a platform that mixes the advantages of mobile health care and big data analysis. Creating because the primary objective _after the information analysis_ the exploration and extraction of data. And observation in real time patients and embrace patient as an active player.

3. INTELLIGENT HEALTH CARE SYSTEM

A. System Characteristic

Our proposition may be a conception of an intelligent platform which we tend to know as Intelligent health care System (IHCS) that may create analysis of big health care data with a fast approach and in an exceedingly real time, data want coming from numerous sources and considerations patients, disease (risk factor), treatments, and doctors, when this analysis the system offer predicted data that reflects the patient's scenario within the future.

1. The system will be hosted in a cloud and might be accessed anytime and anyplace,
2. The system can create a fast analysis in time period to provide accurate future data victimization intelligent and extremely specific tools,
3. The doctor will access to the applying for consulting and observation the health standing, additionally build action if necessary,
4. The patient can still in contact with the intelligent system through an application that will be put in on his Smartphone or on a wise watch with a sensor for detection every modification of his health status (Fig1)

B. Intelligent health care System architecture

IHCS and throughout the medical data can capable to:

- Analyze a large quantity of medical data
- Predict what the patient might have within the future as quality and pathologies by data processing technics
- Anticipate the cure and treatment,
- observation patient in real time ,
- The patient can have the opportunity to create a self-monitoring in time period by the employment of health mobile devices.

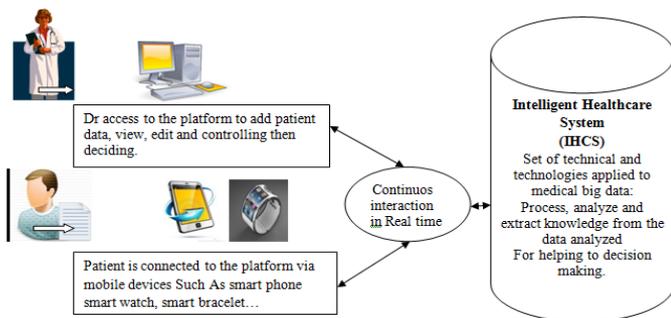


Fig 1. A New schema of intelligent Healthcare system

And to confirm that we tend to build our proposed system and design that combines many steps:

- Data collection: is that the most significant and sensitive part and since the data is that the main element and also the pivot of the system. We should mention that additional data is accurate additional the accuracy of the predicted data will increase.
- The voluminous medical data may be coming from numerous Electronic Health Record (EHR) / Patient Health Record (PHR), Clinical systems and external sources like government sources, laboratories, pharmacies, insurance corporations etc., in varied formats (flat files, .csv, tables, ASCII/text, etc.) .
- Information Warehousing: during this section, massive data returning from various sources warehoused to be cleaned, accumulated and completed for any process.
- Big data analysis: it's an awfully vital part seen it demands very powerful techniques and tools to manage and method the voluminous data.
- Predictive analysis is that the master step altogether this method as a result of it rests on the exploration of analyzed data to extract helpful data on the premise of information mining tools and algorithms to seek out links between the medical data.
- Processing analyzed reports: The results obtained when the predictive analysis process are exploited by:

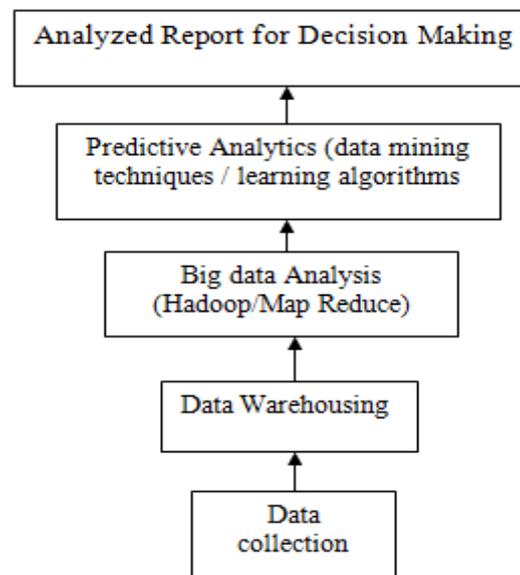


Fig 2. Architecture of the predictive analysis system- Health Care Application

The doctor for facilitate in decision making and giving a general view of the patient's status the patient can have the results of this method by his doctor however he's continuously in interaction with the system by a mobile device that he owned.

In the initial layer Hadoop is employed as an open source framework designed to perform process on massive medical data, The operational principle is as follows, the infrastructure applies the well-known principle of grid computing, of dividing the execution of a method on multiple nodes or clusters of servers. In Hadoop design logic, this list is split into many elements, every half being keep on a unique server cluster. rather than lean process during a single cluster, as is that the case for traditional design, the distribution of data helps distribute the process across all calculate nodes on that the list is distributed. To implement such a technical method, Hadoop is coupled to a file system known as HDFS (Hadoop Distributed file system for).

It manages the allocation of storage of user data in blocks of data on different nodes. HDFS was inspired by a technology utilized by Google to have these cloud services, and known as Google file system (GFS). Map/Reduce: the distribution and management of the calculations is administered by Map reduce. This technology combines two kinds of function:

- The Map function: that resides on the master node then divides the input data or task into smaller subtasks, which it then distributes to employee nodes that method the smaller tasks and pass the answers back to the master node .The subtasks are run in parallel on multiple computers.
- The reduce function: collects the results of all the subtasks and combines them to provide an aggregated final result — that it returns because the answer to the original big question.

The second layer is characterized by the good role of Map-Reduce module for the method of predictive analysis. And to reinforce more and more the system in matters of prediction, it should be equipped by a strong predictive algorithm or learning algorithm to ensure the vital phases of the method and build an appropriate model of prediction.

Data Mining technology sort of a delicate process , executed by prophetic algorithms, that have shown a robust effectiveness and efficiency in predicting , take as an example supporting victor machine (SVM) [31], call tree(C4.5) [32] , and Naive Thomas Bayes (NB) [33], as

they're presently classified Among the highest 10 classification strategies known by IEEE Python & related Resources [34]. For that our system ought to be equipped with a learning algorithms among the cited ones or a mixture of many learning algorithms to learn from its performances and build a robust hybrid algorithm which will be apply to all types of medical prediction.

IV. CONCLUSION

Big data analytics in health care provides all health care delivery system benefits like explorations Data and knowledge extraction, economical value reduction and push medical care to the better, however to offer the facility to predict who can die or who can suffer from diseases and anticipate cure for saving human life, that's a big challenge wish necessitates great efforts to make the mandatory tools and platforms. Our paper projected a design of platform primarily based health care predictive analytics want reply to this request. All that push into the combination and development of predictive analysis algorithm for analyzing medical data as a result of nowadays an easy analyze isn't sufficient contact extracting useful data from it's the biggest challenge.

REFERENCES

- [1] Borkar VR, Carey MJ, Chen L: Big data platforms: what's next? ACM Crossroads 2012, 19(1):44–49.
- [2] Zikopoulos PC, Eaton C, DeRoos D, Deutsch T, Lapis G: Understanding Big Data – Analytics for Enterprise Class Hadoop and Streaming Data. McGraw-Hill: Aspen Institute; 2012.
- [3] N., Cristianini, J, Shawe-Taylor.: An Introduction to Support Vector Machines and Other Kernel-based Learning Methods,Cambridge University Press, 2013, 200 pp, availableat<http://ebooks.cambridge.org/ebook.jsf?bid=CBO9780511801389>
- [4] R. M. Rahman, F. Rabbi: Using and comparing different decision tree classification techniques for mining ICDDR, B Hospital Surveillance data”, elsevier, V, 38, pp 11421–11436.
- [5] S. L. Ang, H. Ch. Ong and H. Chin Low.: Classification Using the General Bayesian Network, science and technology, 24 (1) 205 – 211 pp, (2016)]
- [6] Top Data Mining Algorithms Identified by IEEE & RelatedPythonResources:<http://www.datasciencecentral.co>

m/profiles/blogs/python-
resourcesfortopdataminingalgorithms.

[7] H.Madhusudhana Rao ; H.Madhusudhana Rao ; Dr. B Rambhupal Reddy ; SURVEY OF ADAPTING CLOUD COMPUTING IN HEALTHCARE ; International Journal of Advanced Research in Engineering and Applied Sciences ; 2278-6252

[8] BARBARA CALABRESE; MARIO CANNATARO; CLOUD COMPUTING IN HEALTHCARE AND BIOMEDICIN, SCPE; Volume 16, Number 1, pp. 1–18.

[9] Sanjay P, Ahuja, Sindhu Mani, Jesus Zambrano ; A Survey of the State of Cloud Computing in Healthcare ; Canadian Center of Science and Education. Vol. 1, No. 2; 2012; September 19, 2012; <http://dx.doi.org/10.5539/nct.v1n2p12>

[10] Michael J. Ward, Keith A. Marsolo , Craig M. Froehle . Applications of business analytics in healthcare.ELSEVIER . Business Horizons (2014) 57, 571—582.

[11] Fabricio F. Costa; Big data in biomedicine; ELSEVIER. Drug Discovery Today _ Volume 19, Number 4 _April 2014.

[12] Sunkara . Vaishnavi,Akula.Suneetha ; Role of Cloud Computing in Health Monitoring System ; IJSEAT, Vol 2, Issue 10, October – 2014

[13] Richard Hillestad, James Bigelow, Anthony Bower, Federico Girosi, Robin Meili, Richard Scoville, and Roger Taylor ; Can ElectronicMedical Record Systems Transform Health Care? Potential Health Benefits, Savings, and Costs. Health Affairs, 24, no.5 (2005):1103-1117.

[14] Sungyoung Oh, Jieun Cha, Architecture Design of Healthcare Software-as-a- Service Platform for Cloud-Based Clinical Decision Support Service; Healthc Inform Res. 2015 April; 21(2):102-110. pISSN 2093-3681 • eISSN 2093-369X.

[15] Bonnie Feldman ;Ellen M. Martin ;Tobi Skotnes ; Big Data in Healthcare Hype and Hope ; Business Development for Digital Health ; October 2012. Availabl at content/uploads/2013/02/107279699-Big-Data-in-Healthcare-Hype-and-Hope1.pdf

[16] Srinath Srinivasa ; Sameep Mehta ; Big Data Analytics. Springer ;December 20-23, 2014 ; Volume 8883 2014 .

[17] Dylan Maltby; Big Data Analytics; ASIST 2011, October 9-13, 2011,

[18] Peter Groves; Basel Kayyali ;The ‘big data’ revolution in healthcare ; McKinsy and Company april 2013. Center for US Health System Reform Business Technology Office. Available at <http://digitalstrategy.nl/wp-content/uploads/E2-2013.04-The-big-data-revolution-in-US-health-care-Accelerating-value-and-innovation.pdf>

[19] Steven S. Eisenberg, Practical Predictive Analytics for Healthcare 101.Avanade; 2013

[20] Hsinchun Chen ; BUSINESS INTELLIGENCE AND ANALYTICS: FROM BIG DATA TO BIG IMPACT ; MIS Quarterly

[21] Alfredo Cuzzocrea ; Il-Yeol Song ; Analytics over Large-Scale Multidimensional Data: The Big Data Revolution ; DOLAP’11, October 28, 2011, ACM 978-1-4503-0963-9/11/10.

[22] Rajeev Gupta, Himanshu Gupta, Cloud Computing and Big Data Analytics: What Is New from Databases Perspective?; SPRINGER ; BDA 2012, LNCS 7678, pp. 42–61, 2012.

[23] Nura Esfandiari , Mohammad Reza Babavalian ; Knowledge discovery in medicine: Current issue and future trend ;ELSEVIER ; Expert Systems with Applications 41 (2014) 4434–4463.

[24] Jonathan Northover, Brian McVeigh, Sharat Krishnagiri .Healthcare in the cloud: the opportunity and the challenge. MLD.

[25] Santosh Kumar, Wendy J. Nilsen ; Mobile Health Technology Evaluation ; ELSEVIER ; Volume 45, Issue 2, August 2013, Pages 228–236

[26] Gabriel I. Barbas, Sherry A. Glied, New Technology and Health Care Costs — The Case of Robot-Assisted Surgery; The NEW ENGLAND JOURNAL of MEDICINE. N Engl J Med 2010; 363:701-704August 19, 2010DOI: 10.1056/NEJMp1006602.

[27] Kenney Ng , Amol Ghoting ; PARAMO: A PARAllel predictive MOdeling platform for healthcare analytic research using electronic health records ; ELSEVIER. Journal of Biomedical Informatics 48 (2014) 160–170.

[28] Lena Griebel, Hans-Ulrich Prokosch. A scoping review of cloud computing in Healthcare; BMC Medical Informatics and Decision Making (2015) ; DOI 10.1186/s12911-015-0145-7.

[29] Marianthi Theoharidou, Nikos Tsalis, Smart Home Solutions for Healthcare: Privacy in Ubiquitous Computing Infrastructures.

[30] Larry J. Kricka , Tracey G. Polsky. The future of laboratory medicine — A 2014 perspective.ELSEVIER. Clinica Chimica Acta 438 (2015) 284–303.

[31] Karthik Kambatlaa, Giorgos Kollias. Trends in big data analytics. J. Parallel Distrib. Comput. 74 (2014) 2561–2573.

[32] Vladimir Stantchev,1 Ricardo Colomo-Palacios, Cloud Computing Based Systems for Healthcare. Volume 2014, Article ID 692619, 2 pages.

[33] Alex Guazzeli. Predictive analytics in healthcare The importance of open standards. ZEMENTIS IBM Developers Worker. 29 November 2011. Available at <http://zementis.com/knowledge-base/predictive-analytics-in-healthcare-the-importance-of-open-standards/>

BIOGRAPHIES



Dr.A.R.Pon Periassamy, completed his Ph.D in AUG 2011 in Bharthidasan University, He is working as an Associate Professor in the Department of Computer science. His area of specializations are Digital Principles,Data Mining, artificial intelligence,artificial neural network, image processing and ontology.