p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 04 Issue 05
April 2017

A Big Data Framework Approach in Healthcare Industry

K.Rajesh Khanna[1], Dr.D.Suresh Babu [2], Dr. Vaibhav Bansal[3]
[1] Research Scholar, Department of Computer science, OPJS University, Rajasthan.
[2]Professor & Head, Computer Science Department, Kakatiya Government College, Warangal.
[3]Associate Professor, Department of Computer science, OPJS University, Rajasthan.

ABSTRACT: It has provided tools to accumulate, manage, analyze, and assimilate large volumes of disparate, structured, and unstructured dataproduced by current healthcare systems. Big data analytics has been recently applied towards aiding the process of care deliveryand disease exploration. However, the adoption rate and research development in this space is still hindered by some fundamentalproblems inherent within the big data paradigm. In this paper, we discuss some of these major challenges with a focus on threeupcoming and promising areas of medical research: image, signal, and genomics based analytics.

KEYWORDS-Data Acquisition, Big data analytics, Data Storage and Retrieval.

I. INTRODUCTION

Although there is already a huge amount of healthcare data around the world and while it is growingat an exponential rate, nearly all of the data is stored in individual silos. Data collected by a GP clinicor by a hospital is mostly kept within the boundaries of the healthcare provider. Moreover, data storedwithin a hospital is hardly ever integrated across multiple IT systems. For example, if we consider allthe available data at a hospital from a single patient’s perspective, information about the patient willexist in the EMR system, laboratory, imaging system and prescription databases. Informationdescribing which doctors and nurses attended to the specific patient will also exist. However, in thevast majority of cases, every data source mentioned here is stored in separate silos. Thus derivinginsights and therefore value from the aggregation of these data sets is not possible at this stage. It isalso important to realize that in today’s world a patient’s medical data does not only reside within theboundaries of a healthcare provider. The medical insurance and pharmaceuticals industries also holdinformation about specific claims and the characteristics of prescribed drugs respectively. Increasingly,patient-generated data from IoT devices such as fitness trackers, blood pressure monitors andweighing scales are also providing critical information about the day-to-day lifestyle characteristics ofan individual. Insights derived from such data generated by the linking among EMR data, vital data,laboratory data, medication information, symptoms (to mention some of these) and their aggregation,even more with doctor notes, patient discharge letters, patient diaries, medical publications, namelylinking structured with unstructured data, can be crucial to design coaching programmes that wouldhelp improve peoples’ lifestyles and eventually reduce incidences of chronic disease, medication andhospitalization.

As the healthcare sector transitions from a volume to value-based care model, it is essential fordifferent stakeholders to get a complete and accurate understanding of treatment trajectories ofspecific patient populations. The only way to achieve this is to be able to aggregate the disparate datasources not just within a single hospital’s/GP clinic’s IT infrastructure, but also across multiplehealthcare providers, other healthcare players (e.g. insurance &pharma) and even consumergenerated data. Such unified data sets would benefit not only every player within the healthcareindustry (thus allowing better quality care and access to healthcare at lower costs), but would alsomost importantly benefit the patient by providing first time right treatment, based on a sustainablepricing model.

However, achieving such a vision which involves the integration of such disparate healthcare datasets(in terms of data granularity, quality, type (e.g. ranging from free text, images, (streaming) sensor datato structured datasets) poses major legal, business and technical challenges from a data perspective,in terms of the volume, variety, veracity and velocity of the data sets. The only way to successfullyaddress these challenges is to utilise Big Data technologies.“Big data" has a wide range of definitions in health research1314. However, a viable definition of whatbig data means for health is the following: "Big data in health" encompasses high volume, high diversitybiological, clinical, environmental, and lifestyle information collected from single individuals to largecohorts, in relation to their health and wellness status, at one or several time points. More generaldefinition of Big Data, refers to "datasets whose size is beyond the ability of typical database softwaretools to capture, store, manage and analyse". (McKinsey Global Institute). This definition puts theaccent on the size/volume aspect but, as we described above, the dimensions are many: variety(handling with a multiplicity of types, sources and format), data veracity (related to the quality andvalidity of these data), and data velocity (availability in real time). In addition, there are other factorsthat should also be considered such as data trustworthiness, data protection, and privacy (due to thesensitivity of data managed). All these aspects lead to the need for new algorithms, techniques andapproaches to handle these new challenges.

II. SUGGESTEDWORK

Streaming data analytics in healthcare may be definedas a scientific use of continuous waveform (signal varyingin opposition to time) and related medical record statistics evolved via applied analytical disciplines (e.g., statistical,quantitative, contextual, cognitive, and predictive) to pressurechoice making for patient care. Te analytics workflowof actual-time streaming waveforms in scientific settings canbe broadly defined the usage of Figure 1. Firstly, a platform forstreaming data acquisition and ingestion is needed whichhas the bandwidth to address multiple waveforms at different fidelities. Integrating those dynamic waveform facts withstatic statistics from the EHR is a key component to providesituational and contextual attention for the analytics engine.Enriching the data ate up by way of analytics not most effective makesthe device more strong, but also helps balance the sensitivityand specifcity of the predictive analytics. Te specifics of thesignal processing will in large part depend upon the form of sicknesscohort underneath investigation. A sort of signal processingmechanisms can be applied to extract a mess of targetcapabilities which can be then fed on by way of a pretrained machinemastering version to supply an actionable insight. These actionable insights may want to both be diagnostic, predictive,or prescriptive. These insights may want to in addition be designed tocause other mechanisms together with alarms and notification tophysicians.

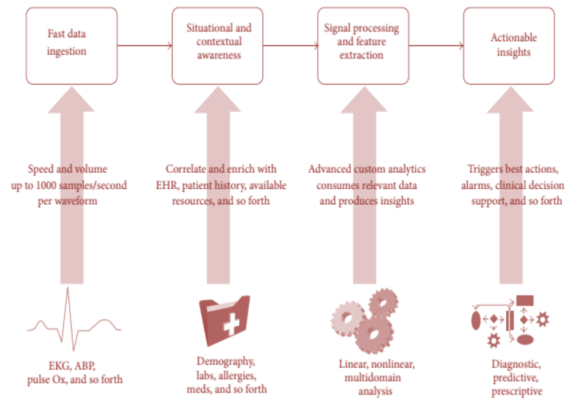

Figure 1: Generalized analytic workflow using streaming healthcare data

Harmonizing such non-stop waveform statistics with discreate records from different resources for locating necessary affected personstatistics and conducting research in the direction of developmentof next technology diagnoses and remedies may be adaunting mission. For bed-side implementation of suchsystems in scientific environments, there are numerous technicalconcerns and requirements that want to be designedand implemented at device, analytic, and scientific ranges.Te following subsections provide a top level view of differentchallenges and current processes within the development ofmonitoring systems that devour both excessive constancy waveform data and discrete data from noncontinuous assets.

Data Acquisition. Historically streaming data from continuous physiological signal acquisition devices was rarelystored. Even if the option to store this data were available, thelength of these data aptures was typically short and downloaded only using proprietary sofware and data formatsprovided by the device manufacturers. Although most majormedical device manufactures are now taking steps to provideinterfaces to access live streaming data from their devices,such data in motion very quickly poses archetypal big datachallenges. Te fact that there are also governance challengessuch as lack of data protocols, lack of data standards, anddata privacy issues is adding to this. On the other side thereare many challenges within the healthcare systems such asnetwork bandwidth, scalability, and cost that have stalled thewidespread adoption of such streaming data collection. This has allowed way for system-wide projects whichespecially cater to medical research communities.Research community has interest in consuming data captured from live monitors for developing continuous monitoring technologies. There have been several indigenousand off-the-shelf efforts in developing and implementingsystems that enable such data capture. There arealso products being developed in the industry that facilitatedevice manufacturer agnostic data acquisition from patientmonitors across healthcare systems.

Data Storage and Retrieval. With large volumes ofstreaming data and other patient information that can begathered from clinical settings, sophisticated storage mechanisms of such data are imperative. Since storing and retrieving can be computational and time expensive, it is key to havea storage infrastructure that facilitates rapid data pull andcommits based on analytic demands.With its capability to store and compute large volumesof data, usage of systems such as Hadoop, MapReduce, andMongoDB is becoming much more common withthe healthcare research communities. MongoDB is a freecross-platform document-oriented database which eschewstraditional table-based relational database. Typically eachhealth system has its own custom relational database schemasand data models which inhibit interoperability of healthcaredata for multi-institutional data sharing or research studies.Furthermore, given the nature of traditional databases integrating data of different types such as streaming waveforms.

This is where MongoDB and other document-based databases can provide highperformance, high availability, and easy scalability for thehealthcare data needs. Apache Hadoop is an opensource framework that allows for the distributed processingof large datasets across clusters of computers using simpleprogramming models. It is a highly scalable platform whichprovides a variety of computing modules such as MapReduceand Spark. For performing analytics on continuous telemetrywaveforms, a module like Spark is especially useful sinceit provides capabilities to ingest and compute on streamingdata along with machine learning and graphing tools. Suchtechnologies allow researchers to utilize data for both realtime as well as retrospective analysis, with the end goal totranslate scientifc discovery into applications for clinicalsettings in an effective manner.

IV. CONCLUSION

Big data analytics which leverages legions of disparate,structured, and unstructured data sources is going to playa vital role in how healthcare is practiced in the future.One can already see a spectrum of analytics being utilized,aiding in the decision making and performance of healthcarepersonnel and patients. Here we focused on three areas ofinterest: medical image analysis, physiological signal processing, and genomic data processing. Te exponential growth ofthe volume of medical images forces computational scientiststo come up with innovative solutions to process this largevolume of data in tractable timescales. Te trend of adoptionof computational systems for physiological signal processingfrom both research and practicing medical professionalsis growing steadily with the development of some veryimaginative and incredible systems that help save lives.

REFERENCES

[1] C. F. Mackenzie, P. Hu, A. Sen et al., “Automatic pre-hospitalvital signs waveform and trend data capture flls quality management, triage and outcome prediction gaps,” AMIA AnnualSymposium Proceedings, vol. 2008, pp. 318–322, 2008.

[2] M. Bodo, T. Settle, J. Royal, E. Lombardini, E. Sawyer, and S.W. Rothwell, “Multimodal noninvasive monitoring of sof tissuewound healing,” Journal of Clinical Monitoring and Computing,vol. 27, no. 6, pp. 677–688, 2013.

[3] P. Hu, S. M. Galvagno Jr., A. Sen et al., “Identifcation of dynamicprehospital changes with continuous vital signs acquisition,” AirMedical Journal, vol. 33, no. 1, pp. 27–33, 2014.

[4] D. Apiletti, E. Baralis, G. Bruno, and T. Cerquitelli, “Real-timeanalysis of physiological data to support medical applications,”IEEE Transactions on Information Technology in Biomedicine,vol. 13, no. 3, pp. 313–321, 2009.

[5] J. Chen, E. Dougherty, S. S. Demir, C. P. Friedman, C. S. Li,and S. Wong, “Grand challenges for multimodal bio-medicalsystems,” IEEE Circuits and Systems Magazine, vol. 5, no. 2, pp.46–52, 2005.

[6] N. Menachemi, A. Chukmaitov, C. Saunders, and R. G. Brooks,“Hospital quality of care: does information technology matter?Te relationship between information technology adoption andquality of care,” Health Care Management Review, vol. 33, no. 1,pp. 51–59, 2008.

[7] C. M. DesRoches, E. G. Campbell, S. R. Rao et al., “Electronichealth records in ambulatory care—a national survey of physicians,” Te New England Journal of Medicine, vol. 359, no. 1, pp.50–60, 2008.

[8] J. S. McCullough, M. Casey, I. Moscovice, and S. Prasad, “Teeffect of health information technology on quality in U.S.hospitals,” Health Affairs, vol. 29, no. 4, pp. 647–654, 2010.

[9] J. M. Blum, H. Joo, H. Lee, and M. Saeed, “Design andimplementation of a hospital wide waveform capture system,”Journal of Clinical Monitoring and Computing, vol. 29, no. 3, pp.359–362, 2015.

[10] D. Freeman, “Te future of patient monitoring,” Health Management Technology, vol. 30, no. 12, article 26, 2009.

[11] B. Muhsin and A. Sampath, “Systems and methods for storing,analyzing, retrieving and displaying streaming medical data,”US Patent 8,310,336, 2012.

[12] D. Malan, T. Fulford-Jones, M. Welsh, and S. Moulton, “Codeblue: an ad hoc sensor network infrastructure for emergencymedical care,” in Proceedings of the International Workshop onWearable and Implantable Body Sensor Networks, vol. 5, London,UK, 2004.

[13] A. Page, O. Kocabas, S. Ames, M. Venkitasubramaniam, andT. Soyata, “Cloud-based secure health monitoring: optimizingfully-homomorphic encryption for streaming algorithms,” inProceedings of the IEEE Globecom Workshops (GC Wkshps ’14),pp. 48–52, IEEE, Austin, Tex, USA, December 2014.

[14] J. Bange, M. Gryzwa, K. Hoyme, D. C. Johnson, J. LaLonde,and W. Mass, “Medical data transport over wireless life criticalnetwork,” US Patent 7,978,062, 2011.

[15] N. Kara and O. A. Dragoi, “Reasoning with contextual datain telehealth applications,” in Proceedings of the 3rd IEEEInternational Conference on Wireless and Mobile Computing,Networking and Communications (WiMoB ’07), p. 69, IEEE,October 2007.

[16] G. Li, J. Liu, X. Li, L. Lin, and R. Wei, “A multiple biomedicalsignals synchronous acquisition circuit based on over-samplingand shaped signal for the application of the ubiquitous healthcare,” Circuits, Systems, and Signal Processing, vol. 33, no. 10, pp.3003–3017, 2014.

[17] A. Bar-Or, J. Healey, L. Kontothanassis, and J. M. van Tong,“BioStream: a system architecture for real-time processingof physiological signals,” in Proceedings of the 26th AnnualInternational Conference of the IEEE Engineering in Medicineand Biology Society (EMBC’ 04), vol. 2, pp. 3101–3104, September2004.

[18] W. Raghupathi and V. Raghupathi, “Big data analytics inhealthcare: promise and potential,” Health Information Scienceand Systems, vol. 2, article 3, 2014.

[19] S. Ahmad, T. Ramsay, L. Huebsch et al., “Continuous multiparameter heart rate variability analysis heralds onset of sepsisin adults,” PLoS ONE, vol. 4, no. 8, Article ID e6642, 2009.

[20] A. L. Goldberger, L. A. Amaral, L. Glass et al., “Physiobank,physiotoolkit, and physionet components of a new researchresource for complex physiologic signals,” Circulation, vol. 101,no. 23, pp. e215–e220, 2000.

[21] E. J. Siachalou, I. K. Kitsas, K. J. Panoulas et al., “ICASP:an intensive-care acquisition and signal

processing integratedframework," Journal of Medical Systems, vol. 29, no. 6, pp. 633–646, 2005.

[22] M. Saeed, C. Lieu, G. Raber, and R. G. Mark, "Mimic ii: amassive temporal icu patient database to support research inintelligent patient monitoring," in Proceedings of the Computersin Cardiology, pp. 641–644, IEEE, September 2002.

[23] A. Burykin, T. Peck, and T. G. Buchman, "Using 'off-theshelf ' tools for terabyte-scale waveform recording in intensivecare: computer system design, database description and lessonslearned," Computer Methods and Programs in Biomedicine, vol.103, no. 3, pp. 151–160, 2011.

[24] G. Adrian, G. E. Francisco, M. Marcela, A. Baum, L. Daniel, and ´G. B. de Quiros Fern ´ an, "Mongodb: an open source alternative ´for HL7-CDA clinical documents management," in Proceedingsof the Open Source International Conference (CISL '13), BuenosAires, Argentina, 2013.

[25] K. Kaur and R. Rani, "Managing data in healthcare informationsystems: many models, one solution," Computer, vol. 48, no. 3,pp. 52–59, 2015.