International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 4 Issue: 10 134 - 138

Big Data in Health Care Analytics

Rohil Shah
Student, Department of Computer Engineering, Dwarkadas J. Sanghvi College of Engineering
Mumbai, India
rgshah3@gmail.com

Ria Echhpal
Student, Department of Computer Engineering, Dwarkadas J. Sanghvi College of Engineering
Mumbai, India
riae1795@gmail.com

Sindhu Nair
Asst. Professor, Department of Computer Engineering, Dwarkadas J. Sanghvi College of Engineering
Mumbai, India
sindhu.nair@djsce.ac.in

***Abstract*—**In the 21st century, availability of large data makes the task of handling it even more difficult. Big data serves to tackle these problems. The growth of data in recent years has been exponential, therefore we explore the prospects, challenges and applications of Big Data. Digitization of Health Care data has seen health records, grown enormously. A patient's medical insurance data, DNA data, medical test results and health history all stored electronically. This data, being unstructured make it very difficult to extract information and analyze patterns that can be useful. We examine the ways in which we can use Big data in health care so that it's potential can be fully tapped.

Keywords*-*big data; healthcare; health; data analytics; applications of big data; prospects of big data.

*****

I. INTRODUCTION

Healthcare, in today's world is at a very crucial stage with a Triple Aim [1], that is better, cheaper and safer healthcare that offers improved outcomes with an efficient and speedy treatment. As the world's population is on the rise, everyone has the desire to live longer and avoid preventable deaths and with desire came the advent of machines. Digitization of Health Care has seen health records grown enormously. Every day, there is enormous amount of data being generated through machines such as ECGs, pathology labs, x-rays and other kind of imagine and signaling machines, but this data is of a very complex and sizeable format. We have to deal with increased patients, more data and very less time. Normal processing on this kind of data is not possible because this data is unstructured and retrieving useful information and analyzing this becomes very difficult. This is where Big Data comes due to its capacity to handle voluminous data which not only has high velocity and variety that is valuable and can be used to improve the health care.

II. BACKGROUND

A. A Brief History about Big Data

The term 'Big Data' was first used in 1997 by Michael Cox and David Ellsworth in an IEEE conference in a paper titled "Application-controlled demand paging for out-of-core visualization". It is the first paper in the ACM to use the term 'Big Data'.

In 1997 Micheal Lesk published "How much information is there in the world?" which concluded that there may be a few thousand petabytes of information and by the year 2000, the production of disks and tapes will reach the level that all the information can be saved and nothing will be thrown out.

In October 1998, K.G. Coffman and Andrew Odlyzko published a paper that concluded the growth of traffic on the Internet was 100% higher than traffic on other networks.

In August 1999, Steve Bryson, David Kenwright, Michael Cox, David Ellsworth, and Robert Haimes publish "Visually exploring gigabyte data sets in real time" in the Communications of the ACM. It is the first CACM article to use the term "Big Data". This article with the opening statement: The article opens with the following statement: "Very powerful computers are a blessing to many fields of inquiry. They are also a curse; fast computations spew out massive amounts of data. Where megabyte data sets were once considered large, we now find data sets from individual simulations in the 300GB range. But understanding the data resulting from high-end computations is a significant endeavor. As more than one scientist has put it, it is just plain difficult to look at all the numbers. And as Richard W. Hamming, mathematician and pioneer computer scientist, pointed out, the purpose of computing is insight, not numbers." [9].

At the beginning of 2009, big data entered the revolutionary stage [8]. Americans consumed information at an average of almost 12 hours a day. Consumption totaled 3.6 Zettabytes and 10,845 trillion words corresponding to 100,500 words and 34 gigabytes for an average person on an average day [9].

The amount of data usage is rising exponentially every year and an estimated total of 2.5 quintillion terabytes of data were generated every day in 2012 alone, and it is estimated that as much data is now generated in just two days as was created from the dawn of civilization until 2003 [6].

B. What Exactly is Big Data?

We currently live in a world that is driven by data which will make a critical difference in our ability to compete in the future. Data is generated from everything around us; every digital process and social media activity is responsible in the production of this kind of data. The reasons behind big data being in the forefront are:

- To find competitive advantages.
- Drives innovation.
- Affects all circles of business.

Big Data has been adopted in many fields and industries today where it has a lot of importance. Places where Big Data is used:

- Understanding and targeting potential customers in a store by taking surveys.
- In financial trading by analyzing the market trends and the history of the companies.

- Car dealerships can take a more competitive stance with customers who prefer a car of a rival company by analyzing customer preferences and product details.

C. The Common Misconception

The biggest misconception about big data is the size of the data. People consider huge terabytes of data as Big Data, which is not entirely true. Small volume of data can also be considered as Big Data since it is the kind of data that cannot be easily stored and processed within a given timeframe. Big data not only defines the size but also finds insights from unstructured, complex, noisy, heterogeneous, longitudinal and voluminous data [2]. Availability of large data makes the task of handling it more difficult. Big data serves to tackle these problems. The growth of data in recent years has been exponential, we therefore need novel methods to handle it.

D. Prospects

Every business is related to data in the present day. The amount of data in the current world is going to exponentially increase and managing it is essential. The methods of analyzing data will also have to improve in this situation. There are a number of industries which are adopting big data and finding ways to implement it.

For instance, there is a new proposed method of squid fishing that makes use of shared information. According to the researchers, "Based on the idea that this commonly used device should be utilizable for the automatic measurement of water temperature profiles, we aimed to develop a ubiquitous sensor that can meet this criterion while also being easily used by fishermen at sea. Through 32 field trials, we confirmed that our device can repeatedly measure the water temperature distribution down to a depth of 300 m. Also, through direct comparison with oceanographic instruments mounted on actual research vessels, we were able to verify that the measurements produced by our device at thermoclines varied by no more than 1.3 °C from the values produced by such instruments." [11]. The ski resorts in Vail uses RFID tags to read and validate 1000 ski passes carried by the guests. While high-frequency (HF) passive RFID tags operating at 13.56 MHz have become standard for ski pass applications, Vail is utilizing newer, ultrahigh-frequency (UHF) passive EPC Gen 2 tags, which operate at 900 MHz and can be read from much greater distances than HF tags. [12] Thus there is a huge room for growth and advancement of big data in all industries, and given a suitable situation, it can do wonders.

E. Applications

Big data has a universal appeal in the industry today. There exist a number of challenges while using Big Data. With that said, a Gartner Survey for 2015 shows that more than 75% of companies are investing or are planning to invest in big data in the next two years. These findings represent a significant increase from a similar survey done in 2012 which indicated that 58% of companies invested or were planning to invest in big data within the next 2 years [10].

- Big Data for Education: In the field of education, Big Data is becoming increasingly hot topic. Educational Data Mining and Learning Analytics are two growing fields of study, trying to make sense of education data and to improve teaching and learning experience [4]. For example, With the help of big data, the teachers can know the extent of knowledge the students have and base their lessons such that it will be optimal for the former to understand.
- Big Data for Media analysis: During the shooting of a film, a large amount of data is generated such as the unused footage, camera information, software and technologies used. The main cinema specific contributions, tested on a multi-source production data set made publicly available for research purposes, are the monitoring and quality assurance of multi camera setups, multisource registration and acceleration of 3-D reconstruction, anthropocentric visual analysis techniques for semantic content annotation, and integrated 2-D/3-D web visualization tools [7]. This available data can be used for future projects to enhance the product and increase the efficiency of production.
- Banking and Security: Big Data can be used by the Securities Exchange Commission to monitor trades and catch illegal trading activity in markets. It can also be used for fraud mitigation, and to obtain information about customers.

III. Big Data in Health Care

Using Big Data Analytics can help solve the spending problem in healthcare. Big Data Analytics can be applied in the diagnosis and treatment of diseases and can reduce the costs that the patients need to pay. The time a doctor allots to every patient has reduced drastically over the years, from about 60 minutes per patient to just over 10 minutes per patient. It is therefore critical that the doctors have all the patients' information in a structured format so that they can easily access this information instead of having the patient bring previous files and the doctor tried to find which prescription was recommended for curing which illness. Thus, the doctor typically has the increasing burden of patient number coupled with less time to spend on each patient coupled with an increasing amount and variety. So we are dealing with more patients, more data, and less time [17].

Although profit should not be the motivation behind this, it is necessary for healthcare organizations to acquire the techniques and infrastructure to leverage big data productively or else risk millions of dollars in profits.

Nowadays, data can be captured from fitness devices, social media and other sources, but very little of this data can be actually captured and organized so that it can be used to provide useful information. Innovative methods are required to capture data, and convert the structured data into unstructured data. Big data in healthcare's true potential lies in merging traditional data with new forms of data to provide faster and more reliable research and discovery.

By digitizing, combining effectively using big data, healthcare organizations can improve their quality of service by analyzing the effectiveness of a treatment and also the efficiency of the healthcare delivery process by detecting fraud, waste and drug abuse more quickly and efficiently [18].

A. The 7 V's of Big Data

Volume: The scale definitely makes Big Data so huge, and this volume is discussed using unimaginable terms with Gigabytes being the past and the recent being Zettabytes or even Yottabytes. To put such volume into context, every minute, 48 hours of video are uploaded to YouTube. The Health Care sector generates data from different sources that is very big in size and needs to be stored in such a way that it is

easily accessible and can help the doctors offer the patients a speedy recovery.

Velocity: Velocity if the time required to access a file. It is needed when the data needs to be stored quickly and refers to the ascending speed at which the data can be accessed and stored into databases. Data can be gathered from everywhere be it hospitals, social media or research facilities, and the most important thing that matters is that speed at which the doctors receive this information and help treat patients at any location.

Variety: The data being structured or unstructured can make variety the biggest challenge of Big Data. Organizing such data in a significant way is very tedious, especially with such a huge volume of rapidly changing data. The complexity of variety comes from the fact that there is no universally accepted approach to handle such data. The data received from ECGs, lab reports, medical history, etc. are all different kinds of data either structured or unstructured and all this data along with experiences during treating a person should be stored together for later use.

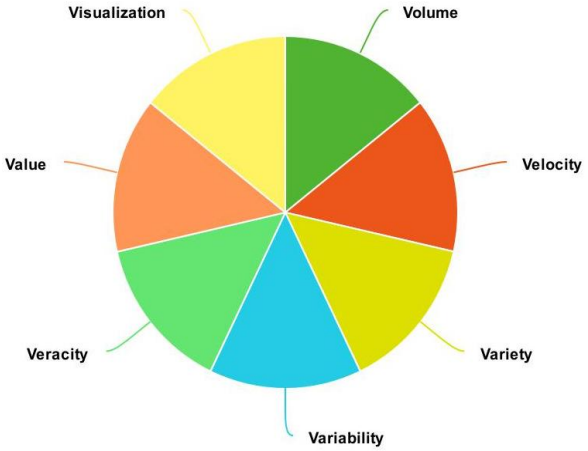

Figure 1. The 7 V's of Big Data

Variability: Variability makes reference to the data whose meaning is repeatedly changing. This data may be the same kind of data, but its value will be different for different circumstances. A single word with multiple meanings is the perfect example to depict this, since new meanings are created, and the old discarded constantly. Understanding this kind of data although challenging is not completely impossible. Every ECG reading holds a different value for a different case may it be a heart attack or a coma. In a similar fashion, patient history can be interpreted in various ways depending on the illness the patient is facing.

Veracity: Veracity deals with the accuracy of Big data, because useless data is completely worthless and misleading, especially in the case of automated machines that use unsupervised machine learning algorithms to make decisions. The findings made in medical research or the data related to cure a disease should be accurate, otherwise it can cause serious issues, leading to the death of the patient.

Visualization: Another challenge of Big data is to gather and transform voluminous data into data that is more easy to comprehend and understand. Once processed, the data needs to be converted into a more readable form that is achieved with the help of graphs and charts. Hospitals are starting to use graph analytics to evaluate the relationship across many complex variables such as laboratory results, nursing notes, patient family history, diagnoses, medications, and patient surveys to identify patients who may be at risk of an adverse outcome [6]. As we can see in Fig 2, a graph is created to depict the frequency of common diseases in all the months of the Year 2015, which is a comprehensible form of representing the data so everyone can read it.

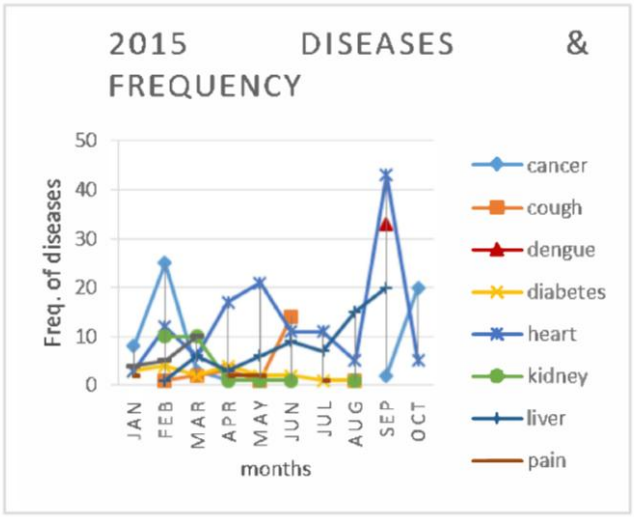

Figure 2. Diseases and their frequency [5]

Value: Having addressing volume, velocity, variety, variability, veracity, and visualization, which takes a lot of time, effort and resources, you want to be sure your organization is getting value from the data [13]. Data in and of itself has no value. The data has value if it is used correctly and for the right purpose. Patient history, previous cases and medical research has no value if it is not used to efficiently treat patients and thus help reduce costs.

B. Levels of Healthcare

Healthcare as a whole can be provided by various entities depending on the criticality and availability of services present in a particular area. It is not definite and is disparately spread over different parts of the world.

As seen in the diagram above, there are three basic levels of healthcare, namely primary, secondary and tertiary. To provide a basic idea, a broad idea for the basis of their classification is as follows:

Primary Healthcare: This is the first level of contact between a member of the community and the healthcare system. It can include paraprofessionals, midwives, and general practitioners. It can also be something as basic as providing education about food and nutrition and the importance of water.

Secondary Healthcare: This is provided by physicians who have basic health education and do not have first contact with the patient. They are usually visited by patients after a referral from a primary health care provider. They may be led by consultant services, such as orthopedics, psychology and psychiatry. In terms of block level secondary health care, it may include District hospitals and Health Care centers at community level.

Tertiary Healthcare: This is the highest level, which provides advanced health care and treatment for patients who have been referred by the primary and secondary levels. This includes treatment and care for intensive and complex medical problems. It includes for example, neurosurgery, cardiothoracic surgery and transplants. It is provided by hospitals and specialized healthcare centers.

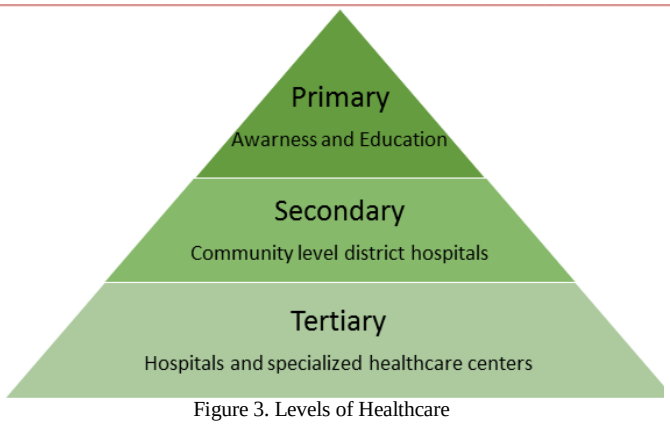

Figure 3. Levels of Healthcare

C. *Sources of Healthcare*

Big data in health-care refers to the patient's data such as physician notes, lab reports, x-ray reports, case history, diet regime, list of doctors and nurses in a particular hospital [l4], and drug analysis, social media, genome research, transactional data etc. [15].

Data, in today's world, can be collected from almost anything, things we can't even imagine, so is the case with data collection in Healthcare. The various sources of Data collection sources are mentioned in Fig y. The data from these sources, being unstructured increases the complexity making it difficult to work with. Therefore, we must use techniques to simplify this data and make it more readable with the help of Hadoop and MapReduce.

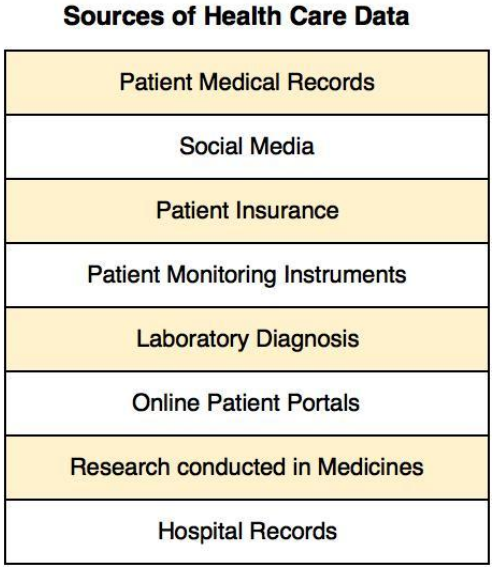

Figure 4. Sources of Health Care Data

D. *Applications of Big Data in Healthcare*

Personalized Health Care: In case of Big Data for Health Care, it can be used to provide an individualistic approach. Using the previous data of a patient, the symptoms of their future illness can be mapped to the most likely outcome. The number of tests that would be required would be less thus saving time of the patient, money of the patient as well as resources of the hospital. In order to illustrate this concept, we will use as a guiding example the problem of predicting the risk of bone fracture in a woman affected by osteoporosis, a pathological reduction of her bone mineralized mass [19]. Big Data can be used to predict the possibility of the woman having a fracture during a certain time period (for example ten years). If a fracture occurs, then this gives the true value taken for prediction in the future.

Healthcare based on Genetics: This means that you use genetic information of a person to determine whether they are susceptible to a particular disease. The DNA of a person is an indication of their genetic composition. By studying this, and the genetic composition of their progenitors, the probability of a disorder being passed can be determined. If the symptoms that a patient has are coherent with certain genetically transferred diseases, Big Data can help to identify the probability of the person having the problem. For example, if a person is showing signs of Alzheimer's, his/her genetic history can be checked to determine whether this is a possibility and the necessary tests can be carried out.

Rare Disorders: In cases of rare disorders, such as endemic diseases that are synonymous to a particular area, or disorders that are found in a very small population, it can be difficult to detect them. In case the symptoms presented by a patient cannot be mapped to any disorder, then as shown in the flowchart, the worldwide data can be checked. This can not only save time and money, but it can also prevent an endemic from turning into an epidemic. Figure 5. Shows how this is done.

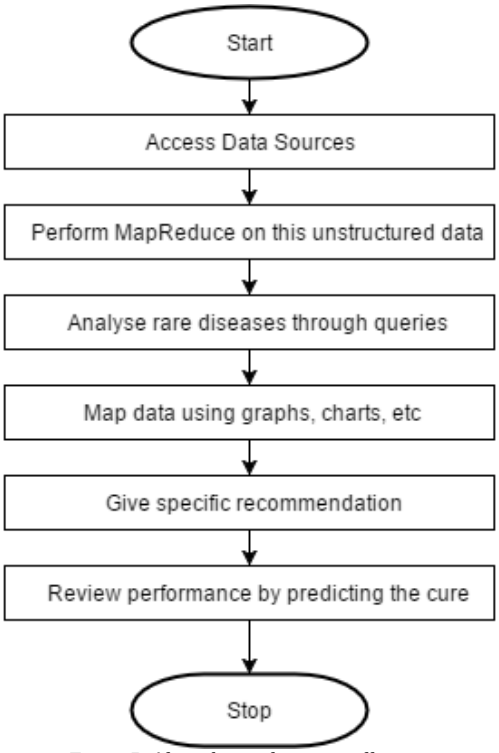

Figure 5. Algorithm to detect rare illnesses

IV. CHALLENGES

There are a number of challenges that come with storage, processing and classifying a set of given data.

The first and most rampant challenge is the speed of browsing and processing the data. In today's industry, where time is of utmost importance, we cannot afford to pause till the required data is retrieved. The challenge only grows as the degree of granularity increases [3].

To overcome this, we can use parallel processing or better hardware so that the large amount of data can be scanned quickly.

- Another challenge is the understanding of the data that has been obtained. The context of the data and the

purpose for which it has to be used should be clear before using it.

- Also, there is a large amount of data available from various sources. The credibility of this data should be checked before using it to reach any conclusion. Visualization is only important if the information provided by the data is relevant and dependable.
- The manner in which the result of the analysis should be displayed in a manner that it is easily understandable by the user. For example, if there is a dataset of 10 billion tweets and it has to be plotted onto a graph with one point for each tweet, it should be clustered and represented such that the result is easily interpreted by the user.
- There have been several developments made towards ensuring that the aforementioned problems do not serve as a hindrance to the evolution of big data, but more efficient technologies need to be developed so that Big Data can be used in more miscellaneous applications.

V. CONCLUSION

As we have seen, Big Data can play a huge role in healthcare but there is a lot of scope for improvement. The cost of using Big Data should be reduced as well as safety and privacy of the data, which is a major concern at the moment should be improved. Big Data can also be used for Insurance reimbursement models wherein the correct ways to fairly recompense doctors can be determined. Also patient satisfaction matrix is another possibility so that the experiences of the patients can help to further improve the system. Besides this, Big data also has scope in staffing and planning, drug and vaccine development and risk identification and mitigation. The possibilities are endless and exploring them will cause mammoth changes in the existing system.

REFERENCES

[1] Institute for Healthcare Improvement: http://www.ihi.org/engage/initiatives/tripleaim/pages/default.aspx

[2] Bernard Marr, how Big data is changing healthcare: https://wtvox.com/big-data/the-future-of-big-data/

[3] Five challenges of big data: https://www.sas.com/resources/asset/five-big-data-challenges-article.pdf

[4] J. Liang, J. Yang, Y. Wu, C. Li and L. Zheng, "Big Data Application in Education: Dropout Prediction in Edx MOOCs," 2016 IEEE Second International Conference on Multimedia Big Data (BigMM), Taipei, 2016, pp. 440-443.

[5] A. K. Bamwal, G. K. Choudhary, R. Swamim, A. Kedia, S. Goswami and A. K. Das, "Application of twitter in health care sector for India," 2016 3rd International Conference on Recent Advances in Information Technology (RAIT), Dhanbad, 2016, pp. 172-176.

[6] https://hbr.org/2014/12/why-health-care-may-finally-be-ready-for-big-data

[7] S Haykin; V Tresp; J.A. Benediktsson.2016.Proceedings of the IEEE,Year: 2016, Volume: 104, Issue: 11Pages: 2082 - 2084

[8] Bryant, R.E., Katz, R.H., and Lazowska, E.D. 2008. BigData Computing: Creating revolutionary breakthroughs in commerce, science, and society Computing, In Computing Research Initiatives for the 21st Century, Computing Research Association, available at http://www.cra.org/ccc/files/docs/init/Big_Data.pdf.

[9] A Very Short History Of Big Data: http://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/#6f370feb55da

[10] Applications of Big Data in industries: https://www.simplilearn.com/big-data-applications-in-industries-article

[11] M. Wada, "Prospects of the future squid fishing based on the big data," OCEANS 2015 - MTS/IEEE Washington, Washington, DC, 2015, pp. 1-6.

[12] Vail picks new line with UHF RFID-Powered Passes: http://www.rfidjournal.com/articles/view?4193

[13] The 7 V's of Big Data: https://www.impactradius.com/blog/7-vs-big-data/

[14] Kapil Khandelwal "Is „Big Data" Big Business in Healthcare in India", Mentor, Investor and a Healthcare Expert, the article originally Appeared in the print version of Health Biz India in May 2013.

[15] M. Ojha and K. Mathur, "Proposed application of big data analytics in healthcare at Maharaja Yeshwantrao Hospital," 2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC), Muscat, 2016, pp. 1-7. doi: 10.1109/ICBDSC.2016.7460340

[16] F. Rahman and M. J. Slepian, "Application of big-data in healthcare analytics - Prospects and challenges," 2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), Las Vegas, NV, 2016, pp. 13-16.

[17] R. Sathiyavathi, A Survey: Big Data Analytics on Healthcare System, Contemporary Engineering Sciences, Vol. 8, no. 3, 121 - 125 HIKARI Ltd, www.m-hikari.com, 2015.

[18] M. Viceconti, P. Hunter and R. Hose, "Big Data, Big Knowledge: Big Data for Personalized Healthcare," in IEEE Journal of Biomedical and Health Informatics, vol. 19, no. 4, pp. 1209-1215, July 2015. doi: 10.1109/JBHI.2015.24068