Big Data Application for Breast Cancer Treatment

Hady Sylla
Indiana University
Smith Research Center
Bloomington, IN 47408, USA
hsylla@iu.edu

ABSTRACT

This paper is about Big Data application for the treatment of breast cancer.The paper explore contribution that big data had on the analysis of numerous data by researchers

KEYWORDS

HID 339, Big Data, Breast Cancer, Cancer

1 INTRODUCTION

We are living in the time of Big Data, which has encouraged a significant discovery in medicine and will support most, if not all, of the treatment and anticipation progresses yet to come[7].The International Cancer Genome Consortium. The Cancer Genome Atlas, and by international consortia, e.g., Information inside any one asset can go from a specialty informational index to completely coordinated information created by numerous innovation stages and a large number of patient examples[7].In the past, scientists were aware of the fact that cancers frequently have extra or missing chromosomes or pieces of chromosomes, which is referred to as aneuploidy

2 BIG DATA APPLICATION FOR BREAST CANCER TREATMENT

There are 3 billion DNA code letters in each human cell and 32 thousand billion cells in the body. So every individual has 96 thousand billion DNA code letters[4]. That is more than ten times the quantity of code letters as there are grains of sand in most of the shorelines on Earth. Besides, a lamentable difference in any of those code letters can begin a sickness or make it impenetrable to medical research[4].

The term "breast cancer" encompasses many different cancers as no two breast cancers are precisely the same. Researchers utilize genomic technology to describe these cancers fully and develop applications of this knowledge that guide treatment decisions tailored to fit the needs of individual patients[1].But, it was it was unclear until recently if this characteristic was significant or merely the byproduct of tumor growth[9].In 2013. The research conducted by geneticist Stephen Elledge identified aneuploidy as the factor that is responsible for driving cancer[9].This discovery was derived using tremendous amounts of cellular data and the ability of computers to aid researchers in sifting through this information[9].

3 BIG DATA TO ADVANCE BREAST CANCER RISK PREDICTION

Researchers supported by Cancer Research have made a 'guide' connecting the state of the city with of breast cancer cells to genes switched on and off, and coordinated it to genuine malady results, as indicated by an examination distributed in Genome Research[6]. Big-data researchers utilize an extensive data set, such as the Cancer Genome Atlas (TCGA), and look for patterns within the data[1].The goal is to identify mutations, which researchers can target by drug treatment that they personalized to a particular patient's needs[1].Big data research involves analyzing the data derived from thousands of tumors, which reveal patterns that can improve screenings and diagnosis, as well as guide treatment[1].[5]provide an overview of data resources on cancer-related research. These authors review compendia of data resources, a list of cancer-related data resources, and a biomolecular repository "'Hubs,'" as well as lists of seminal publications and journals on data science[5].In other words, this review describes where researchers can find breast cancer data and aids determining the range of data types that are available[5].

When a cancer patient is diagnosed, the tumor's genome can be sequenced, and this information can be used to identify drugs that are likely to affect tumor growth (Savage, 2014). Elledge's discovery that aneuploidy is the engine driving cancer growth resulted directly from a computational method developed by his researcher and his colleagues, the Tumor Suppressor and Oncogene Explorer (TSOE)[9].The TSOE is used to mind large data sets, which include the Cancer Genome Atlas and the Catalogue of Somatic Mutations in Cancer[9].There were roughly 70 suppressor genes and 50 oncogenes already known, but the development of the TSOE increased these number to approximately 329 and 200 respectively[9].Analytical data is available on 8,200 tumors, but researchers consider this to be just a start[9].

[11]report that rapid, yet accurate, text-mining using empirical literature makes possible the discovery of new knowledge that will help researchers obtain a better understanding of human diseases, which can then be used to improve the care delivery. These researchers designed and developed a text-mining framework that they refer to as Spark-Text, which utilizes a "Big Data infrastructure" that includes "Apache Spark data streaming and machine learning methods, combined with a Cassandra NoSQL database"[11].The researcher extracted information relevant to several types of cancers, including breast cancer, accessing tens of thousands of articles. The researchers conclude that the potential for mining scientific articles

using this Dig Data infrastructure is very high[11].Furthermore, the SparkText program can be utilized in other areas of biomedical research[11].

[9]points out that a significant problem with the vast data sets relevant to genomic cancer data based on biomarkers concerns developing methods for manipulating this information, which can terabyte level and beyond. Big data infrastructures, such as the one developed by[11].offers means for utilizing this invaluable data and using this information to inform screening, diagnosis, and delivery of quality care to patients. Individually, big-data science has led to researchers rethinking how to breast cancer[1].

While mining big data holds the potential of leading to a medical breakthrough, the information is analyzed thoroughly, and it is also necessary to understand the pros and cons of big data analytics[2].The advantages include the fact that big data focused on correlations, not causality, which means that big data sets have the power to alert researchers to patterns that they did not expect [2].Big data allows healthcare providers to personalized treatment to fit the needs of individual patients. A University of Ontario study of sepsis in premature babies demonstrates how analysis of large data set can provide correlations that lead to clinical actions [2].By employing the data from 1200 data points-per-second, generated from wireless sensors attached to babies, the researchers successfully diagnosed infections 24 hours before fever development and increases in white blood cell count [2].

In another research study distributed by Genome Research, the researchers effectively mapped the state of bosom tumor cells to qualities, and coordinated it to illness results[3]. This guide could help doctors in picking a treatment for patients. The researchers utilized extensive datasets to establish a ink between cell shape and qualities. Generally, they inspected more than 300,000 bosom tumor cells and 28,000 distinct qualities[3]. The investigators found that NF-kappaB is a central protein involved with the network, and could promote proliferation and metastasis of cancer cells. This was linked to cancer stage, and may be used to predict survival outcomes in patients with breast cancer. Through big data approach, the analysts could filter through a huge number of disease cells and qualities to decide their affiliations. This guide could be utilized by doctors later on to decide the treatment alternative that has prompted the most elevated survival rate in patients with comparable malignancies. It could likewise give understanding to both the patient and the doctor about the idea of the ailment.

Furhermore, Madabhushi worked with Shannon C. Agner at Rutgers University and Mark A. Rosen, MD; Sarah Englander; Mitchell D. Schnall, MD; Michael D. Feldman, MD; Paul Zhang, MD; and Carolyn Miles; MD, at the University of Pennsylvania, on the breast cancer study. They broke down MR pictures of bosom injuries from 65 ladies. The specialists filtered through several gigabytes of picture information from every patient to attempt to discover contrasts that recognize the diverse breast cancer subtypes. The researchers scientifically demonstrated the surfaces that show up as the tissues retain differentiate improving color. The model uncovered that progressions over just milliseconds recognized triple-negative from kind sores. The examiners utilized machine learning and example acknowledgment techniques to help in analyze among the three sorts of growths in view of surface changes and other quantitative proof[10]. Madabhushi posited that Today, if a lady or her specialist finds a protuberance, she gets a mammogram and after that a biopsy for atomic examination, which can take two weeks or up to a month. In the event that we can anticipate the malignancy is triple-negative, we can quick track the patient for biopsy and treatment. Particularly in cases with triple-negative malignancy, two to a month spared can be pivotal[10].

A discussion by Clifford Hudis, MD, at the fifteenth St Gallen Breast Cancer Conference stated that The lack of patients taking part in clinical trials makes information extrapolation and application complex, requiring a need to investigate wellbeing innovation arrangements that tap the capability of genuine information[8]. In the United States, just roughly 3% of patients determined to have malignancy are enlisted in clinical trials. Besides, significant difference exists in socioeconomics between members in clinical trials and the all inclusive community. Regardless of electronic record selection, "one impediment to curing malignancy stayed: tolerant information isn't shared," said Hudis. Indeed, wellbeing data got under this demonstration was not interoperable, speaking to a noteworthy obstacle. As a rule, the electronic wellbeing records were fundamentally kept up with the goal that clinicians could enough guard their charging hones for expensive medications and therapeutics upon review. The answer for this issue was a framework called CancerLinQ, noted Hudis, who is the present seat of the huge information activity's governing body. CancerLinQ incorporates quiet information, inside privacy rules, and takes into consideration information mining and sharing. "The basic role of CancerLinQ is to enhance the nature of care and to upgrade results," Hudis stated[8]. By March 2017, almost 2 million records had been joined in the framework from 80 oncology mind settings, which extended from singular practices to huge growth focuses. The framework benefits from the information as of now being entered. In a normal day, around 40% of a clinicians time is currently spent on record entering, as per Hudis. Hudis conclude by stating that traditional research drives us forward yet is restricted by a tight pool of subjects, and the cost and long time expected to build up a result. Later on, huge informational collections may expand and broaden the fantastic confirmation from planned research that incorporates more seasoned patients, comorbidities, simultaneous medicine and numerous other certifiable ramifications of fruitful tumor treatment[8].

4 CONCLUSION

As another field of research, the look for measures that boost predictivity may do much in the method of satisfying the expectations of progressing anticipating results of intrigue. Big data allows the dissecting information from numerous disease sorts that researchers can assess prognostic models and recognize quality changes that prompted tumor formation.

ACKNOWLEDGMENTS

I would like to thank Dr. Gregor von Laszewski for his support and suggestions to write this paper.

REFERENCES

[1] Jill U Adams. 2015. Genetics: big hopes for big data. *Nature* 527, 7578 (2015), S108–S109.

[2] Kent Bottles and Edmon Begoli. 2014. Understanding the pros and cons of big data analytics. *Physician executive* 40, 4 (2014), 6.

[3] Leo Breiman. 1996. Bagging predictors. *Machine learning* 24, 2 (1996), 123–140.

[4] Luigi Luca Cavalli-Sforza and Walter Fred Bodmer. 1999. *The genetics of human populations*. Courier Corporation.

[5] Susan E Clare and Pamela L Shaw. 2016. fiBig Datafi for Breast Cancer: Where to look and what you will find. *NPJ breast cancer* 2 (2016).

[6] Robert A Hiatt and Barbara K Rimer. 1999. A new strategy for cancer control research. *Cancer Epidemiology and Prevention Biomarkers* 8, 11 (1999), 957–964.

[7] Travis B Murdoch and Allan S Detsky. 2013. The inevitable application of big data to health care. *Jama* 309, 13 (2013), 1351–1352.

[8] Virginia Powers. 2017. Big Data May Guide Future Treatment Decisions. *onclive* (2017). http://www.onclive.com/conference-coverage/st-gallen-2017/big-data-may-guide-future-treatment-decisions

[9] Neil Savage. 2014. Bioinformatics: big data versus the big C. *Nature* 509, 7502 (2014), S66–S67.

[10] Samuel Fosso Wamba, Shahriar Akter, Andrew Edwards, Geoffrey Chopin, and Denis Gnanzou. 2015. How fibig datafican make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics* 165 (2015), 234–246.

[11] Zhan Ye, Ahmad P Tafti, Karen Y He, Kai Wang, and Max M He. 2016. Sparktext: Biomedical text mining on big data framework. *PloS one* 11, 9 (2016), e0162721.

@Articlemurdoch2013inevitable, author = Murdoch, Travis B and Detsky, Allan S, title = The inevitable application of big data to health care, journal = Jama, year = 2013, volume = 309, number = 13, pages = 1351–1352, publisher = American Medical Association,

@Articleadams, author = Adams, Jill U, title = Genetics: big hopes for big data, journal = Nature, year = 2015, volume = 527, number = 7578, pages = S108–S109, publisher = Nature Research,

@Articlebottles2014, author = Bottles, Kent and Begoli, Edmon, title = Understanding the pros and cons of big data analytics, journal = Physician executive, year = 2014, volume = 40, number = 4, pages = 6, publisher = American Association for Physician Leadership,

@Articleclare2016, author = Clare, Susan E and Shaw, Pamela L, title = fiBig Datafi for Breast Cancer: Where to look and what you will find, journal = NPJ breast cancer, year = 2016, volume = 2, publisher = NIH Public Access,

@Articlesavage2014, author = Savage, Neil, title = Bioinformatics: big data versus the big C, journal = Nature, year = 2014, volume = 509, number = 7502, pages = S66–S67, publisher = Nature Research,

@Articleye2016, author = Ye, Zhan and Tafti, Ahmad P and He, Karen Y and Wang, Kai and He, Max M, title = Sparktext: Biomedical text mining on big data framework, journal = PloS one, year = 2016, volume = 11, number = 9, pages = e0162721, publisher = Public Library of Science,

@Bookcavalli1999, title = The genetics of human populations, publisher = Courier Corporation, year = 1999, author = Cavalli-Sforza, Luigi Luca and Bodmer, Walter Fred,

@Articlehiatt1999new, author = Hiatt, Robert A and Rimer, Barbara K, title = A new strategy for cancer control research, journal = Cancer Epidemiology and Prevention Biomarkers, year = 1999, volume = 8, number = 11, pages = 957–964, publisher = AACR,

@Articlebreiman1996bagging, author = Breiman, Leo, title = Bagging predictors, journal = Machine learning, year = 1996, volume = 24, number = 2, pages = 123–140, publisher = Springer,

@Articlewamba2015big, author = Wamba, Samuel Fosso and Akter, Shahriar and Edwards, Andrew and Chopin, Geoffrey and Gnanzou, Denis, title = How fibig datafican make big impact: Findings from a systematic review and a longitudinal case study, journal = International Journal of Production Economics, year = 2015, volume = 165, pages = 234–246, publisher = Elsevier,

@Articlevir, author = Virginia Powers, title = Big Data May Guide Future Treatment Decisions, journal = onclive, year = 2017, url = http://www.onclive.com/conference-coverage/st-gallen-2017/big-data-may-guide-future-treatment-decisions,

5 BIBTEX ISSUES

Warning–empty address in cavalli1999

Warning–page numbers missing in both pages and numpages fields in clare2016

Warning–no number and no volume in vir

Warning–page numbers missing in both pages and numpages fields in vir

(There were 4 warnings)

6 ISSUES

DONE:

Example of done item: Once you fix an item, change TODO to DONE

6.1 Formatting

Other formatting issues - references section

6.2 Writing Errors

Errors in title, e.g. capitalization - Big Data Applications...

Do not use the phrase *In this paper/report we show* instead use *We show*. It is not important if this is a paper or a report and does not need to be mentioned

Use a space after . , :

6.3 Citation Issues and Plagiarism

Need to quote directly cited material

6.4 Character Errors

Erroneous use of quotation marks, i.e. use “quotes” , instead of ” ”

6.5 Structural Issues

Acknowledgement section missing