Predictive Analytics: Healthcare Benefits and Consequences

Christine Orosco
Bellevue University
Bellevue, NE 68005, USA
cjorosco@my365.bellevue.edu

Blaine Blasdell
Bellevue University
Bellevue, NE 68005, USA
beblasdell@my365.bellevue.edu

Saima Rahmanzai
Bellevue University
Bellevue, NE 68005, USA
szrahmanzai@my365.bellevue.edu

CHI 2020 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6819-3/20/04.
DOI: https://doi.org/10.1145/3334480.

Abstract

This paper presents a review of the impact Predictive Analytics to population health. More specifically, the review describes the role of Predictive Analytics in three of the top deadliest diseases afflicting Americans today. These are Cancer, Alzheimer's, and Heart Diseases. The objectives of the review are to ascertain the positive impacts and contributions to combating these diseases. Another objective is to identify ethical and privacy issues with the early diagnosis or the prediction of outcome from the diseases. With Electronic Health Records (EHR) being the norm, the likelihood of inadvertent disclosure of this data will happen. Even with the most stringent security controls applied to EHR, this does not stop employers, insurers, or other agencies from gaining access to EHR data. What are the ramifications for disclosing the results from Predictive Analytics recorded in EHR?

Author Keywords

Predictive Analytics; Population Health; Data Science Ethics and Privacy; Disease Predictions.

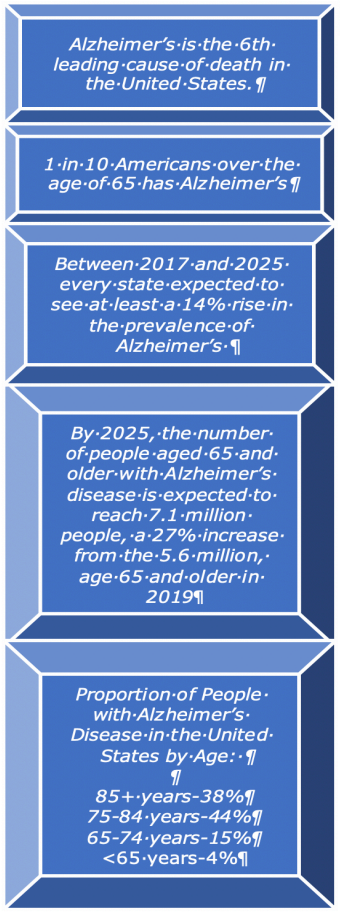

Figure 1 AD in the USA [1]

CSS Concepts

*•**Applied computing**~Life and medical sciences~Consumer health•**Applied computing**~Life and medical sciences~Health informatics•**Security and privacy**~Human and societal aspects of security and privacy*

Introduction and Motivation

This paper addresses the benefits and consequences that Predictive Analytics (PA) may have on three of the top diseases afflicting Americans here in the United States. As shown in Figure 1, Alzheimer's Disease (AD) continues to grow as our population ages [1]. Astonishing figures such as **1 in 10 Americans over the age of 65** has Alzheimer's illustrates the necessity of finding a cure and the cause of Alzheimer's in the near future. Cancer is on the rise as our population grows and ages. Figure 2 [2] lists the types of cancer with the diagnosis and death rates for each type. The American Cancer Society [2] estimates approximately **1.8 million people** will be diagnosed with cancer and approximately **606,000 people will die** from cancer in 2020. Another startling statistic from the CDC is that **1 person dies every 37 seconds** in the US from cardiovascular disease [3]. Approximately **647,000 Americans die** from heart disease each year [3]. To put that number in perspective, that equates to **1 in 4 deaths** are attributed to heart disease [3]. It is the severity of these diseases and the impact to population health that are the impetus for the paper. With the integration of Machine Learning (ML) and Artificial Intelligence (AI) in many aspects of healthcare, an objective of this research is to identify several instances that PA has and proposes to improve the research and

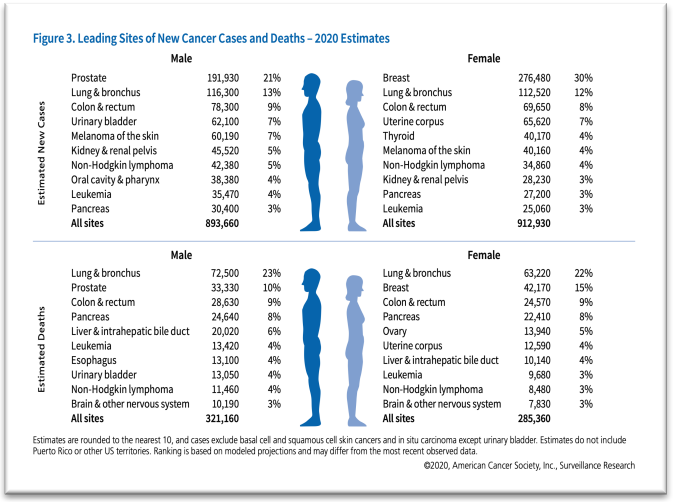
Figure 3. Leading Sites of New Cancer Cases and Deaths – 2020 Estimates

Estimated New Cases

Male			Female		
Prostate	191,930	21%	Breast	276,480	30%
Lung & bronchus	116,300	13%	Lung & bronchus	112,520	12%
Colon & rectum	78,300	9%	Colon & rectum	69,650	8%
Urinary bladder	62,100	7%	Uterine corpus	65,620	7%
Melanoma of the skin	60,190	7%	Thyroid	40,170	4%
Kidney & renal pelvis	45,520	5%	Melanoma of the skin	40,160	4%
Non-Hodgkin lymphoma	42,380	5%	Non-Hodgkin lymphoma	34,860	4%
Oral cavity & pharynx	38,380	4%	Kidney & renal pelvis	28,230	3%
Leukemia	35,470	4%	Pancreas	27,200	3%
Pancreas	30,400	3%	Leukemia	25,060	3%
All sites	**893,660**		**All sites**	**912,930**	

Estimated Deaths

Male			Female		
Lung & bronchus	72,500	23%	Lung & bronchus	63,220	22%
Prostate	33,330	10%	Breast	42,170	15%
Colon & rectum	28,630	9%	Colon & rectum	24,570	9%
Pancreas	24,640	8%	Pancreas	22,410	8%
Liver & intrahepatic bile duct	20,020	6%	Ovary	13,940	5%
Leukemia	13,420	4%	Uterine corpus	12,590	4%
Esophagus	13,100	4%	Liver & intrahepatic bile duct	10,140	4%
Urinary bladder	13,050	4%	Leukemia	9,680	3%
Non-Hodgkin lymphoma	11,460	4%	Non-Hodgkin lymphoma	8,480	3%
Brain & other nervous system	10,190	3%	Brain & other nervous system	7,830	3%
All sites	**321,160**		**All sites**	**285,360**	

Estimates are rounded to the nearest 10, and cases exclude basal cell and squamous cell skin cancers and in situ carcinoma except urinary bladder. Estimates do not include Puerto Rico or other US territories. Ranking is based on modeled projections and may differ from the most recent observed data.

Figure 2 2020 CDC Cancer Estimates [2]

Benefits of Predictive Analytics

Predictive Analytics is beneficial to any industry provided the results and outcomes are applied in ways that benefit not only the industry segment but also humankind. This is especially true for PA in the healthcare segments. Predictive Analytics is not a panacea rather it is a tool to assist in the diagnosis and treatment of many diseases. It supports improvements in workflow management, in turn, benefits the delivery of healthcare services.

Literature Research

The following paragraphs describe studies using PA to identify the propensity for Alzheimer's, Cancer, and Heart disease or assist in the early intervention and treatment of each disease.

Figure 2 2020 CDC Cancer Estimates.

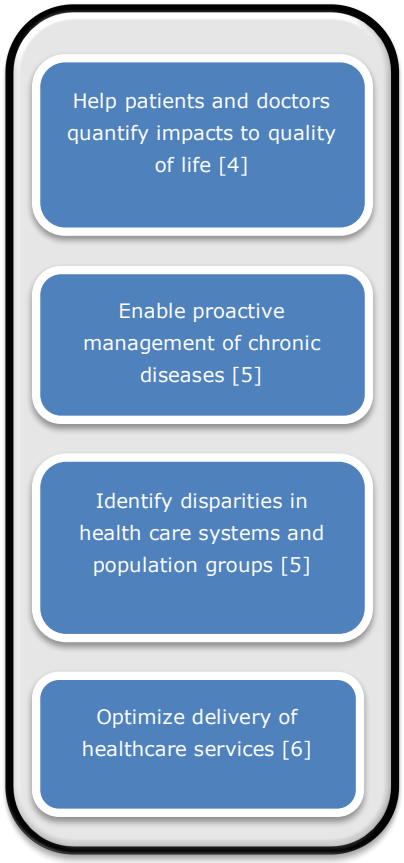

Figure 3 Benefits of PA in Healthcare

Alzheimer's Disease is becoming an epidemic in adults 65 and older. It accounts for over 60-80% of all dementia cases. The early detection of the disease is the focus of the survey in [7]. This survey reviews studies for Alzheimer's research that focus on the early detection of the disease using Support Vector Machine, Logistics Regression, Random Forests (RF), Bayesian Networks and Deep Learning algorithms

In [8] researchers for the first-time used RF algorithms on neuroimaging data and combined this with demographics and genetic data, and cognitive scores. The purpose is to predict the conversion rate of Mild Cognitive Impairment (MCI) to AD. The study uses ML classifiers with RF from the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) system. The authors reviewed 34 articles and selected 12 articles addressing MCI and AD published between 2007 and 2017 for inclusion into their study. The findings of the study using RF, produced some of the highest accuracies (90%) to-date for prediction and early diagnosis of AD [8]. The benefit of using RF classification is that it promotes identification of the features that most contribute to the prediction of AD and MCI by brain region and the possible connections to each pathology.

In a partnership between the National Cancer Institute (NCI) and U.S. Department of Energy (DOE), the two organization are using high performance computing and PA on an Exascale amount of data for three new pilot projects [9]. These three studies are still in development.

Breast cancer in women has a global incidence rate higher than any other cancer. It is responsible for 500,000 deaths annually [10]. This study uses ML to predict the survival rates for patients with triple-negative breast cancer (TNBC) for 5 years after discharge. The survival rate is very low for this type of cancer. Early detection is critical for long term survival of the patient. This study is beneficial in identifying those breast cancer patients that remain a post-surgical high-risk. The survival rate for this group is only 15%. The study's objective is to assist in the application of early and proactive treatments to help reduce breast cancer mortality rates. The study uses 5 algorithms to predict the survival rates for 1570 stage 1-III patients. These algorithms are Logistics, Decision Trees, RF, Gradient Boosting Machine (GBM).

The focus of the study in [11] is to predict Coronary Heart Disease (CHD) for Korean patients. The objective is to discover a more accurate prediction model. The model that the study proposes is called the CHD Prediction Model. It consists of generating rules using the Classification and Regression Tree (CART) algorithm and applying a Fuzzy Logic method to address the uncertainty problem. This method produces the CHD predictions using datasets from the Korean National Health and Nutrition Examination Survey VI (KNHANES-VI). The research shows that the proposed model does the prediction accuracy and sensitivity. The authors attribute the improved scores to the reduction of uncertainty using Fuzzy Logic method in addition to the CART algorithms.

ADA does not prohibit discrimination based on future medical conditions
No law prohibits employers and health insurers from considering non-genetic predictive data
Data brokers could sell medical predictions to any interested third party. HIPAA Rule does not apply to data brokers. They do not have to ask the patient for permission.
Predictions may lead to psychological harm and even contemplate suicide.
There are no counselling services to advise patients regarding the predictive data.
There are no regulations regarding algorithms used in medical predictions and diagnosis. Is the medical practitioner responsible for misdiagnosis from defective algorithms?

Table 4 Predictive Data Consequences [14]

In the last study [12 the authors propose a system using Cloud Computing resources to predict cardio-vascular disease in patients more accurately than on standalone computing platforms. The study has two phases. The first phase is to predict the occurrence of algorithms and use Hadoop and predictions and improve the accuracy rate. The study uses the Support Vector Machine (SVM), Naïve-Bayes, Multi-layer Perceptron (MLP), RF, AdaBoost boosted Trees, and Binary Discriminants algorithms. The results show that Naïve Bayes provides the more accurate results, followed by AdaBoost.

Privacy and Ethical Issues

There is no doubt the integration of Predictive Analytics into mainstream clinical medicine will continue to have tremendous benefits to population health. Yet these advances do come with consequences. There are significant ethical and privacy issues when it comes to predicting the propensity for acquiring a fatal disease. The creators and consumers of PA data need to be aware of the Privacy, Ethical, and Security issues concerning the protection and processing of sensitive data. They must be knowledgeable of the Health Insurance Portability and Accountability Act (HIPAA), Personally Identifiable Information (PII), and other Privacy requirements. The collection and processing of data is largely unregulated in the United States except for specific segments such as healthcare and payment card processing [13]. Additionally, medical researchers need to be aware of the sensitivity of the predictions and outcomes. What if the patient is predicted to?

As Predictive Analytics evolves from the research stage to the clinical implementation stages, the more likely that predictions about a person's health will become part of the patient's health records. Although HIPAA regulates access to medical records, those within the medical community do have access to these records as do those to whom the patient has given consent. As the author in [14] discusses the collection of EHR with other background information may lead to a potential for discrimination. For example, an employer may use chromic health predictions to deny employment. Would these same people be denied insurance coverage due to their life expectancy or future health status? Figure 4, highlights significant issues and consequences of PA to population health.

Conclusion

Predictive Analytics is experiencing unprecedented growth in its applications across all industry segments. Predictive Analytics is advancing health care in many areas to include identifying propensity towards certain diseases, identifying markers or behavioral changes for early diagnosis of diseases, and identifying health impacts to societal segments. Yet these advances do come with consequences. How do we balance the need for advances to benefit society as a whole with the protection of individual privacy and freedom? Are we heading towards a society that is under constant surveillance? It is ironic that Data Science has given us the insight as to what will happen when too much data is collected and analyzed.

Acknowledgements

Bellevue University Data Science Department and Dr. Shankar Parajulee for their development of graduate student knowledge. For all the countless authors for the continuous research and posts on their subject matter. For the consumer privacy and ethics advocates and regulators to keep pushing the data science discipline,

and in particular, predictive and in particular, predictive analytics to address the ethical and privacy issues so the consumers can enjoy its benefits without the risks posed by the issues.

References

[1] Alzheimer's Statistics, 2019 Alzheimer's Statistics, Alzheimer's.net, 2019. Retrieved from https://www.alzheimers.net/resources/alzheimers-statistics/

[2] American Cancer Society, Cancer Facts and Figures 2020, American Cancer Society, Inc, No. 500820. Retrieved from https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2020/cancer-facts-and-figures-2020.pdf

[3] Heart Disease Facts, CDC, Dec 2019. Retrieved from https://www.cdc.gov/heartdisease/facts.htm.

[4] ElieMassaad, Nida Fatima, Muhamed Hadzipasic, Christopher Alvarez-Breckenridge, Ganesh Shankar, Joh Shin, Predictive Analytics in Spine Oncology Research: First Steps, Limitations, and Future Directions, Neurospine 31 Dec. 2019.

[5] Muin Khoury, Michael Engelgau, David Chambers, George Mensah, Beyond Public Health Genomics: Cam Big Data and Predictive Analytics Deliver Precision Public Health?, Public Health Fenomics, 17 Jul. 2019.

[6] M. Malik, S. Abdallah, M. Ala'raj, Data Mining and predictive analytics applications for the delivery of healthcare services: a systematic literature review, Annuals of Operations Research, 24 Dec. 2016.

[7] Rui Zhang, Gyorgy Simon, Fang Yu, Advancing Alzheimer's Research: A Review of Big Data Promises, International Journal of Medical Informatics, October 2017.

[8] Alessia Sarica, Antonio Cerasa, Aldo Quattrone, Random Forest Algorithm for Alzheimer's disease the Classification of Neuroimaging Data in Alzheimer's Disease: A Systematic Review, Frontiers in Aging Neuroscience, Oct. 6 2017.

[9] Esther Shein, Combating Cancer with Data, Communications of the ACM, May 2017, Vol. 60. No.5. pp 10-12.

[10] Yucan Xu, Lingsha Ju, Jianhua Tong, Chengmao Zhou, Jianjun Yang, Supervised Machine Learning Predictive Analytics For Triple-Negative Breast Cancer Death Outcomes, OncoTargets and Therapy, Dec 2019.

[11] S. Vaanathi, Cardiovascular Disease Prediction Using Fuzzy Logic Expert System, IUP Journal of Computer Sciences, Jul 2017, Vol.XI, No. 3, pp.7-21.

[12] Nishant Gupta, Naman Ahuja, Shikhar Malhotra, Anju Bala, Gurleen Kaur, Intelligent Heart Disease Prediction in Cloud Environment through Ensembling, Wiley Expert Systems, Vol. 34, Issue 3, Jun 2017.

[13] Ravi Parikh, Ziad Obermeryer, Amol Navathe, Regulation of Predictive Analytics in Medicine: Algorithms must meet regulatory standards of clinical benefit, Science, 22 Feb 2019.

[14] Sharona Hoffman, Artificial Intelligence in Medicine Raises Legal and Ethical Concerns, The Conversation, 4 Sept. 2019. Retrieved from http://theconversation.com/artificial-intelligence-in-medicine-raises-legal-and-ethical-concerns-122504