2017 IEEE 41st Annual Computer Software and Applications Conference

A Study of Machine Learning in Healthcare

Rohan Bhardwaj
Cisco Systems Inc (and Rice University Houston, TX)
275 East Tasman Drive
San Jose, CA 95134, USA
rb43@rice.edu

Ankita R. Nambiar
Cisco Systems Inc (and Dougherty Valley High School, San Ramon, CA)
275 East Tasman Drive
San Jose, CA 95134, USA
anknambi@cisco.com

Debojyoti Dutta
Cisco Systems Inc.
275 East Tasman Drive, San Jose, CA 95134, USA
dedutta@cisco.com

***Abstract*— In the past few years, there has been significant developments in how machine learning can be used in various industries and research. This paper discusses the potential of utilizing machine learning technologies in healthcare and outlines various industry initiatives using machine learning initiatives in the healthcare sector.**

Keywords—machine learning; healthcare, big data

I. HEALTHCARE IN GENERAL

Healthcare is one of the fastest growing sectors today and is currently in the core of a complete global overhaul and transformation. Russell Reynolds and Associates cites that global healthcare costs, currently estimated at $6 trillion to $7 trillion, are projected to reach more than $12 trillion within just seven years [1]. This trend is also exemplified domestically, in the United States. The total spending on healthcare in the United States increased up to 5.3 percent and has topped $3 trillion nationwide [2]. Additionally, healthcare spending in the United States represents 17 percent of the total gross domestic product (GDP); our healthcare costs are rising at rates close to double of our economic growth rate. In addition to a rise in the amount consumers are spending on healthcare, the federal government has been forced to pay more and more for healthcare as costs become too high for patients to afford. The amount of money the federal government has allocated for healthcare spending has increased by 11.7 percent in 2014 to an incredible $844 billion in 2015[5]. This rise in federal funding represents the significant disparity between the cost of healthcare and the financial burden on consumers[4][7].

Given this rapid growth in costs, a number of actions must be taken to ensure the costs of healthcare do not further spiral out of control. The need for patient-physician communication, follow-up appointments, and the availability of specialists have also become painfully apparent. Innovation and technological solutions may be the solution to fix the issues with our modern-day healthcare system. These innovations range from swallowable microchips that alert doctors when medication has been taken to large scale data analysis to determine which medications are most effective. However, recently, machine learning has been identified as having major technological application in the healthcare realm. While such technologies will probably never completely replace physicians, they can transform the healthcare sector, benefiting both patients and providers.

II. HISTORICAL PERSPECTIVE AND TRENDS

The field of medicine has taken significant strides in its advancement; the development of vaccinations, antibiotics, and even the concept of sterilization have disrupted the industry and caused a cascade-like effect on all patients and doctors involved in healthcare. Needless to say, the human population has progressed a great degree from past medical care. The healthcare industry is made up of preventive, diagnostic, remedial, and therapeutic sectors. Each of these sectors work together to provide a comprehensive, holistic experience for the modern-day patients.

Some major trends are occurring in the industry today; the first of which involves the transition to patient-centered care. Organizational changes have caused a transition from being focused on hospital care to being more reliant on preventative and outpatient centric care. As medicine reaches the apex of its transition to consumerism, there will be an inherent need to provide consumers with the tools to make intelligent decisions [8]. A fundamental aspect of patient orientated care is regarding the strength of the relationship between the physicians and their patients. James Rikert provides the following example, "(A patient's) primary malignancy was lung cancer. During the course of treatment… (she had seen) … a pulmonologist for her symptoms. He had performed pulmonary function tests, prescribed inhalers, and told her to return if her symptoms did not improve. She never went back, and the cancer was later found by her family doctor, by which time it was metastatic" [5]. A fundamental issue in the example above and in healthcare system in the past has been that of not developing a good relationship between the health provider and the patient. The way that a modern-day physician interacts with their patients determines their satisfaction with the care, their response to medication, and their tendency to schedule follow up appointments. In the status quo, doctors have too many patients to ensure that each patient is in good health, even after the appointment. Additionally, doctors are being urged to see a greater number of patients in a smaller amount of time. As the need to establish relationships with patients increases in tandem with the rising number of patients and need for care, doctors will have to utilize new technology to accomplish their goal of patient-centered care [3]. In the example above, technology could have signaled to both the doctor and the

0730-3157/17 $31.00 © 2017 IEEE
DOI 10.1109/COMPSAC.2017.164

patient that there was an apparent need to schedule a follow-up appointment.

III. Healthcare with technologies like Big Data

Big data and analytics have been causing disruptions in major industry segments. In recent years, big data has become a new ubiquitous term. Big Data refers to large, complex datasets that are beyond the capabilities of traditional data management systems of storing, managing, and processing in a timely and economical manner. Big data technologies can handle structured, semi- structured, and unstructured data in petabytes and more. [4][5][6][10].

Healthcare is one of the verticals that can significantly benefit from the increasing amounts of data and its availability [7]. Entities- including health care providers, pharmaceutical companies, research institutions, and government agencies- have begun to compile massive amounts of data from research, clinical trials, and public health and insurance programs. The consolidation of data from various sources has significant potential [5]. In the past, doctors have been treating patients based on symptoms; however, physicians are beginning to diagnose and treat patients with a concept known as evidence-based medicine. This involves reviewing large amounts of data aggregated from clinical trials and other treatment pathways on the large scale and making decisions based on the best information available [8]. For example, if a patient comes in with a particular case of the flu, a physician in the past would rely on what he or she knew about the flu in general or what other doctors in the area knew. With big data technologies, a physician can look at nationwide trends on what course of treatment would work best for the patient to prescribe the best medications. The aggregation of individual data sets that would otherwise prove meaningless provides doctors with the information needed to make better, more holistic medical decisions.

IV. Machine Learning

In today's connected world, data across sectors are growing exponentially. As the volumes of data increase, new novel ways to interact with and to extract meaning from the data are emerging. In the past, human intervention has been used to parse through the data; however, this is inefficient and a large number of hidden patterns within the data are not found [11]. This is externally important is the healthcare sector. As Thomas H. Davenport writes in the Wall Street Journal, "Humans can typically create one or two good models a week (while) machine learning can create thousands of models a week".

Machine Learning is a particular method of data analytics that automates model building, as it relates to the development of models. With machines learning to utilize certain algorithms, they can find hidden insights from data; it is important to note that in machine learning, we are not telling the machines where to look. The iterative nature of machine learning allows the machine to adapt its methods and outputs as it is exposed to new situations and data [12].

There are two main of approaches to go about machine learning. The first one is supervised learning. The realm of supervised learning involves training algorithms using particular examples. The machine receives a number of inputs with a given number of correct outputs; the learning occurs by comparing empirical results with the correct outputs to identify errors [13]. This type of learning is used when past history can be used to predict events in the future [14].

The other approach is unsupervised learning. Under this approach, the machine must explore the data and attempt to develop some sort of pattern or structure; it must develop models from scratch and is not told any correct outputs. This method is often used to pinpoint and distinguish outliers [11].

There are a number of areas currently in which machine learning is heavily being applied. The first of which is in the realm of finance. Supervised learning is used to predict when transactions with credit cards are likely to be fraudulent; these algorithms automatically send a message to credit card holders to check if the charge was indeed fraudulent. Additionally, it is often used in the retail industry. By using predictive analytics powered by machine learning, producers can determine what customers might want next and can create an online experience that caters to the consumers' need [11]. For example, Amazon.com utilizes machine learning to show certain products to customers based on what they have viewed in the past. Example: recommending protein powder after a customer has bought a set of barbells.

V. Machine learning in healthcare

One area, in particular, in which machine learning has possible widespread societal impacts is in the healthcare realm. In a growing industry of smart watches, fit bits, and devices that constantly gather a plethora of health data, the prevalence of using machine learning to analyze this data is gaining momentum [15]. Machine learning can prove to be the solution for both reducing the rising cost of healthcare and helping establish a better patient-doctor relationship. Machine learning and big data solutions can be used for a plethora of health-related uses; some include helping doctors determine more personalized prescriptions and treatments for patients and also helping patients determine when and if they should schedule follow up appointments.

Recently in healthcare, a large amount of data has become available. This includes EMRs that contain data that can be either structured or unstructured [16]. Structured health data is the information that is easy to categorize in a database; they can include a series of statistics and categories including but not limited to patient weights, temperatures, and even generic symptoms like headache, stomach pain etc. [28]. The majority of medical data is unstructured data in the form of various different notes, reports, discharge summaries [15], images, and audio and video recording. It is very hard to quantify and categorize a conversation between the provider and the patient; the conversation is very personalized and can take many different directions [13]. For example, for two patients with the same exact strain of a cold, the conversation

and data would differ depending on the background of the patient, the background of the doctor, and even the different way the patient describes his or her symptoms. In general, structured data makes up about 20 percent of current EMRs; the other 80 percent is made up of unstructured data.

Since the nature of medicine is linked to that of a sort of narrative, modern day machine learning techniques must look towards organizing and establishing a relationship between mass amounts of unstructured, raw data. Being able to harness and understanding this sort of data on a large scale will prove to be extremely beneficial in applying machine learning technologies in the healthcare realm [17]. Currently, a large number of artificial data technology exists for structured data; however, a much smaller percentage of current innovators are focusing on structured data, focusing on the narratives present in the healthcare realm. Machine learning, when applied effectively, can help physicians make near-perfect diagnoses, choose the best medications for their patients, determine patients at high-risk for poor outcomes to medication, and improve patients' general health while reducing cost [13].

As costs for healthcare stagnate at historically high prices and the need for medical oversight increases, machine learning on large scale unstructured data may prove to be the solution to this ever-growing issue. A couple of companies and individuals have established themselves in the market today with their machine learning technology applied to modern medicine with both unstructured data and structured data.

In healthcare, 50 percent of the total costs come from 5 percent of total patients; additionally, the number of chronic conditions requiring consistent, constant care has gradually increased across the country. Machine learning can prove to identify patients who may be more prone to recurring illnesses and help diagnose patients. Additionally, close to 90 percent of emergency room visits are preventable. Machine learning can be used to help diagnose and direct patients to proper treatment all while keeping costs down by keeping patients out of expensive, time intensive emergency care centers [18].

VI. Industry and Research Initiatives

There are a number of industry and research initiatives that aim to apply machine learning technology in the healthcare realm in order to improve the lives of patients around the world [29]. The first of which is a research laboratory based out of Stanford University in Palo Alto, CA. The Shah Lab is headed up by Dr. Nigam Shah and is part of the center of Biomedical Informatics Research at Stanford. The researchers at the Shah Lab use machine learning and data mining in medical ontologies "to enable the learning health system." The researchers look at Electronic Health Records that include longitudinal data regarding patient care—what is important to note is that this initiative analyzes unstructured data [19]. As said before, this type of data will prove key to the overall goal of combining machine learning with the healthcare realm as it represents the narrative nature that makes up medicine.

By looking at patterns for what has been done before in medicine in the large scale, the lab attempts to answer clinical questions, generate data driven insights, and build predicative models regarding the progression of disease, the effectiveness of treatments, and the processes of medical care [19]. All of these insights, the Shah Lab believes, will help doctors make better decisions. Additionally, the lab's insights have impacts in other areas of the healthcare realm. Their pharmacovigilance methods can flag adverse drug events early even before drug recalls. They have also found that androgen deprivation therapy, a treatment for prostate cancer, can actually increase the chance of developing Alzheimer's disease [19]. In 2014, the Shah Lab published a paper stating "Our models are able to predict a future diagnosis of depression up to 12 months in advance." They were able to develop regression-based models for predicting depression, its severity, and its response to treatment based on EMR[30].

Their research shows that machine learning is a very viable option used to interpret large amounts of data, in order to, draw patterns and conclusions that would not otherwise be apparent. It also shows how important machine learning can be to the healthcare industry by aiding doctors in making more informed diseases and helping them distinguish between various treatment options. All of these factor into the vision of the Shah Lab which is to "support clinical decision making by the bedside" [19].

Another initiative that is gaining traction in the healthcare sector is Lumiata, a healthcare startup specialized in graph analytics. Lumiata is headed by Dr. Igor Barani and was recently named by MIT technology review as one of the top 50 smartest companies of 2016 [31]. Enlitic, another startup company, is utilizing deep learning and machine learning to harness a new era of data-driven medicine. "The company's tool aims to augment doctors and make it possible to distill actionable diagnostic insights in real time from millions of prior patient cases and other medical data" [26]. The company believes that at its core, medical diagnostics is a data issue. The key is to turn medical data such as lab x-rays and images, patient histories, and physician notes into meaningful insights and patterns. In a nutshell, it compares medical images to large datasets of past images and data to offer incredibly valuable and accurate insights to physicians. It is important to note that this certain software offers accuracy, efficiency, and transparency. The doctor always understands and is made aware of the reasoning behind all the technical insights. Enlitic technology is truly ground breaking and the company is on the fore front of the effort to combine machine learning in the healthcare realm. Recently, it was found that Enlitic technology detected lung cancer nodules in chest CAT Scan (CT) images 50 percent more accurately as compared to an expert panel of radiologists. This sort of technology can prove to be vital in lowering healthcare costs and helping doctors provide a more personalized, patient centered approach to medicine [23].

Earlier in the paper it was said that the current healthcare system often pressures doctors to see a greater number of patients in a smaller amount of time. Additionally, as the need

to establish relationships with patients increases in tandem with the rising number of patients and need for care, doctors will need to utilize new technology to accomplish their goal of patient-centered care. The company Ginger.io is trying to do just that—to allow patients to have a meaningful, effective relationship with their health care provider in the mental health realm [27]. An online personal coach is in the core of the ginger.io methodology and technology. These coaches supplement health plans with information provided by the ginger.io smartphone app. The human coach has the goal of helping to coordinate patient care to ensure they are getting the treatment they need, when they need it [23].

The coach helps bridge the current gap between the patient and the physician by following up on appointments, scheduling appointments with various therapists, and even providing proactive help. Ginger.io software can detect when the patient has not contacted any contacts in a number of days and sends that information to the coach. Additionally, the software helps to compile, with the help of a physician and coach, a list of medications that might be helpful for the patient. This technology will improve the healthcare realm in a number of ways. First, by inserting a middleman with the time and resources to truly check up on a patient, the physician does not have to be strained for time and resources. Additionally, by utilizing machine learning, health care providers can get valuable information about their patients' lives that would only be apparent from spending more time with them (such as the fact that they are not communicating with friends recently). Ginger.io is unique in that it is utilizing machine learning from a patient viewpoint and not of a physician alone [24].

From a different aspect, the process of getting medical treatments and medication approved for use is a long, arduous task. Many patients are harmed because a medication that could have possibility saved or improved their live is caught in the approval process. The company Berg and Interrogative Biology is attempting to change the medical realm by utilizing machine learning and artificial intelligence in medical research to further develop the idea of personalized treatments for patients [24]. The science all starts with the patient's own biological samples; Berg Interrogative looks at both the activity of healthy and unhealthy cellular matrixes and uses adaptive omic molecular data to study the biological samples. This combination of systems -biology and artificial Intelligence- is known as the "BERG Interrogative Biology® platform." By utilizing the platform's models, scientists are able to go about drug discovery by converting biological output to possible therapeutic candidates and information [24]. The platform collects a large number of data points sampled from patient biology and clinical information. Data collected includes but is not limited to the genome, proteome, lipidome and metabolome, mitochondrial function, oxidative states, and ATP production [25]. All this information helps to paint a sort of portrait of the biological process in a patient. From these network models, novel biological points of interests can be discovered. These points of interest can be considered, in a way, to be biomarkers—they can serve as indicators of diseases[24]. For example, let us look at a cancer patient about to receive treatment for his or her condition. If the patient uses Berg Interrogative, the platform can flag certain drugs for toxicity depending on the patients' respective metabolism, biological makeup, and other information. This process could prove to be extremely helpful in catering treatments to the patients and making sure the treatments will prove to be effective [24].

By tagging molecules that are over- or under-expressed in damaged cells, Interrogative Biology can often find biomarkers to help with the prognosis of a disease. In other words, Berg is applying an AI approach in the search for novel drugs; however, these are not just any drugs—these are natural disease compounds the human body creates to ward off disease. BPM 31510, identified using the latter platform, is Berg's lead molecule, currently in phase I and II trials. This molecule is naturally secreted by the body and was found to restore the metabolism of cancerous cells. This is important because it can prove to trigger the molecular mechanisms that detect cell damage and signal cell death. It is important to note that this molecule is not synthetic and, since it is naturally produced by the human body, will have a much easier process of approval [25]. All of the potential drugs they look at are based upon endogenous enzymes or proteins, The CTO of the company, Niven Narain, suspects that "Berg will have an easier time with toxicology than other pharma companies, because its therapeutics are already present in the human body" [25]. Berg Interrogative is revolutionizing the healthcare industry by helping physicians personalize treatment options to patients and by discovering and testing drugs naturally created by the human body-drugs. These drugs can go into the industry in almost half the time.

Recent years, a number of startup companies have emerged with machine learning and artificial intelligence in the healthcare sector. CB Insights has published a list of 106 such startup in their February 2017 article titles "From Virtual Nurses to Drug Discovery: 106 Artificial Intelligence Startups in Healthcare" [20]. The article quotes Harpreet Singh Buttar from Frost & Sullivan, "By 2025, AI systems could be involved in everything from population health management, to digital avatars capable of answering specific patient queries". Few of the companies from the CB Insights article and extract of their analysis are listed below [20]:

Ayasdi: Ayasdi helps companies use artificial intelligence and big data to make employees more productive and to drive fundamental breakthroughs that are beyond the capabilities of humans. The company's artificial intelligence platform leverages automation, machine learning, and topological data analysis. The data analysis can simplify the extraction of knowledge from large and complex data sets and facilitate the deployment of intelligent, AI-based applications[20].

Apixio: Apixio is a pioneer in applying big data technology to extract and analyze clinically unstructured and coded data using natural language processing and machine learning to derive outcomes and cost-related insights. Apixio offers solutions that enable health plans and large provider

groups to more accurately and efficiently determine member risk scores and receive appropriate value-based reimbursement [20].

BenevolentAI: BenevolentAI was founded in 2013 by Ken Mulvany founder of Proximagen. BenevolentBio is focused on applying technology in the bioscience industries. The initial focus has been on human health – generating new ideas that have the potential to improve the lives of millions and deliver better medicines to patients faster in currently overlooked areas such as orphan disease and rare cancers. BenevolentTech is developing an advanced artificial intelligence platform that helps scientists make new discoveries and redefines how scientists gain access to, and use, all the data available to drive innovation. The technology is built upon a deep judgement system that learns and reasons from the interaction between human reasoning and data[21].

Butterfly Network: Butterfly Network is transforming diagnostic and therapeutic imaging with devices, deep learning, and the cloud. Butterfly Network operates at the intersection of engineering and medicine by bringing together world-class talent in computer science, physics, mechanical engineering, electrical engineering and medicine [20].

Digital Reasoning Systems: Digital Reasoning is a global leader in using artificial intelligence to understand human communications. Its cognitive computing platform, Synthesys, automates key tasks and uncovers transformative insights across vast amounts of human communications for many of the world's most elite companies, organizations, and agencies[20].

Flatiron Health: Flatiron Health is a health care technology company and operator of the OncologyCloud platform. Integrating across the entire clinical data spectrum, Flatiron Health allows cancer care providers and life science companies to gain deep business and clinical intelligence through its web-based platform[20].

H2O.ai: H2O is a provider of an open source based predictive analytics platform for data scientists and application developers who need scalable and fast machine learning for smart business applications. These applications include smart home appliances, self-driving cars, personalized digital content, smart assistants, and others[20].

iCarbonX: CarbonX is developing an artificial intelligence platform to facilitate research related to the treatment of diseases, preventive care, and precision nutrition. This approach is considered an essential element in enabling the future development of personalized medicine [20].

Pathway Genomics: Pathway Genomics, founded in 2008, provides physicians and their patients with accurate genetic information to improve or maintain health and wellness. The company's mobile health applications merge artificial intelligence and deep learning with personal genetic information that provides personalized health and wellness guidance [20].

WellTok: WellTok combines knowledge of the healthcare industry and social networking technology in its CafeWell.com channel to achieve levels of consumer engagement for healthcare population managers through Social Health Management. WellTok's software/Internet products focus on providing a complete, integrated solution that includes the engaging social health network CafeWell; actionable member and group analytics; and integration with enterprise information systems[20].

Another startup in this space is AliveCor - the Silicon Valley-based maker of the Kardia Mobile, a portable electrocardiogram device - is now betting that artificial intelligence will help doctors monitor patients' heart conditions. Its machine learning algorithms will automatically flag abnormal ECGs, leading to early detection of common heart arrhythmias and helping prevent strokes[22].

VII. CONCLUSION

Healthcare is one of the fastest growing sectors in today's economy; more people require care, and it is becoming more and more expensive. Government spending on healthcare has reached an all-time high while the inherent need for enhanced patient-physician connection becomes readily apparent. Technologies like big data and machine learning have the potential to help both patents and providers in terms of better care and lower costs. A number of companies and organization have already taken the first step in this industry and have helped facilitate the transition to patient and evidence-orientated care. The data is there; we just have to figure out how to interpret it—companies like the ones mentioned above are just a small number of the entities taking us one step closer to that vision.

VIII. ACKNOWLEDGEMENT

The authors thank Raghunath Nambiar for his mentorship in writing this paper. The authors also thank Aayush Patel for his reviews.

REFERENCES

[1] Bouwens J. : Embracing Change: The healthcare industry focuses on new growth drivers and leadership requirements
[2] Pianin E.: US Health Care Costs Surge to 17 Percent of GDP. The Financial Times
[3] Brown B. : Top 7 Healthcare Trends and Challenges from Our Financial Expert
[4] Nambiar R., Sethi A., Bhardwaj R., Vargheese R. : A look at challenges and opportunities of Big Data analytics in healthcare. IEEE Big Data Conference 2013
[5]: Kayyali B., Knott D., Kuiken S. : The Big Data Revolution in US Health Care: Accelerating value and innovation. McKinsey & Company
[6] SAS Institute: Big Data: What it is and why it matters?
[7] Bhardwaj R., Adhiraaj Sethi A., Nambiar R. : Big data in genomics: An overview. IEE Big Data Conference 2014
[8] Applod K. : Five big industry changes to watch in 2016
[9] Rickert J. : Patient-Centered Care: What It Means And How To Get There
[10] Daveport T. : Industrial-Strength Analytics with Machine Learning. The Wall Street Journal

[11] SAS Institute: Machine Learning: What it is and what it matters
[12] Wikipedia: Machine Learning
[13] Maddux D. : The Human Condition in Structured and Unstructured Data. Acumen Physician Solutions
[14] Brownlee J. : What is Machine Learning: A Tour of Authoritative Definitions and a Handy One-Liner You Can Use. www. machinelearningmastery.com
[15] Dolley S. : Big Data Solution to Harnessing Unstructured Data in Healthcare.
[16] Hauskrecht M., Visweswaran S., Cooper G., Clermont G.: Clinical Alerting of Unusual Care that Is Based on Machine Learning from Past EMR Data
[17] Page D. : Challenges in Machine Learning from Electronic Health Records. MLHC 2015
[18] Dolley S. : Big Data Solution to Harnessing Unstructured Data in Healthcare. www.cloudera.com
[19] Shah Lab Website: https://shahlab.stanford.edu/
[20] CB Insights: From Virtual Nurses to Drug Discovery: 106 Artificial Intelligence Startups in Healthcare
[21] BenevolentAI website: www.benevolentai.com
[22] AliveCor Website: www.alivecor.com.com
[23] Ginger.io Website: www.ginger.io
[24] Berg Health Website: https://www.berghealth.com
[25] Krol A. : Berg and the Pursuit of the Body's Hidden Drugs
[26] Enlitic Website: www.enlitic.com
[27] Fast Company Staff: The World's Top 10 Most Innovative Companies in Health Care
[28] Kohn M. : Real World Data and Clinical Decision Support
[29] Taghizadeh G. : Top 5 Companies Revolutionizing Healthcare with Machine Learning
[30] Huang SH., LePendu P., Iyer SV., Tai-Seale M., Carrell D., Shah NH. : Toward personalizing treatment for depression: predicting diagnosis and severity.
[31] MIT Technology Review: 50 Smartest Companies 2016