Comprehensive Study on Health Care Diagnosis, Data science and its opportunities.

Venkata Anil Kumar Thota (012525734)

Nandini Puppala(0137527131)

Prathusha Koouri (013710658)

San Jose State University

This paper is part of Mini Project submission for Data296 class (Fall Semester 2018).

Contact: nandini.puppala@sjsu.edu

venkataanilkumar.thota@sjsu.edu

prathusha.koouri@sjsu.com

Contents

Abstract

The rapidly expanding field of data analytics plays a pivotal role in the evolution of healthcare practices and research. It has provided tools to accumulate, manage, analyze, and assimilate large volumes of disparate, structured, and unstructured data produced by current healthcare systems. Data analytics is now being applied towards aiding the process of care delivery and disease exploration. However, the adoption rate and research development in this space is still hindered by some fundamental problems inherent within the data paradigm.

In this paper, we discuss the strategies used by physicians to make diagnoses and the latest techniques used by data scientists to make a prediction based on the supervised learning of the medical data available. We have also discussed jobs, job roles, salary ranges, and startups working on healthcare analytics and tried to analyze the future of an individual working in the field of Data Analytics in Healthcare.

Introduction

The diagnosis of many diseases has become increasingly complex these days. Many different results obtained from tests with substantial imperfections, must be integrated into a diagnostic conclusion about the probability of diseases in a given patient.

To approach this problem in a practical manner, we generally have to review the literature to estimate the pre-test likelihood of disease (defined by age, sex and symptoms) and the sensitivity and specificity of the diagnostic tests. With this information, test results can be analyzed by using the statistical techniques available.

This approach has several advantages. It pools the diagnostic experience of many physicians and integrates fundamental pretest clinical descriptors with many varying test results to summarize reproducibly and meaningfully the probability of a disease. This approach also aids but does not replace the physician’s judgment and may assist in decision making(more accurate results). The diagnosis of a disease on the basis of history and physical examination alone is often difficult. Many sophisticated tests have thus been developed to allow an early and more accurate diagnosis. Although many tests are now firmly established in clinical practice, none is specifically suited to wide-scale, cost-effective application, because each has limitations concerning sensitivity and specificity.

The rapidly expanding field of big data analytics has started to play a pivotal role in the evolution of healthcare practices and research. It has provided tools to accumulate, manage, analyze, and assimilate large volumes of disparate, structured, and unstructured data produced by current healthcare systems. Big data analytics has been recently applied towards aiding the process of care delivery and disease exploration. However, the adoption rate and research development in this space is still hindered by some fundamental problems. Transitioning US healthcare into the digital era is necessary to reduce operational costs at Healthcare Organizations (HCO) and provide better diagnostic tools for Health Professionals.

The strategies used by Physicians

A clinician’s ability to diagnose accurately is central in assessing prognosis and prescribing effective treatments. We found that diagnostic reasoning can be split into a three-stage model:

1)Initiation of diagnostic hypotheses.

2)Refinement of the diagnostic hypotheses.

3)Defining the final diagnosis.

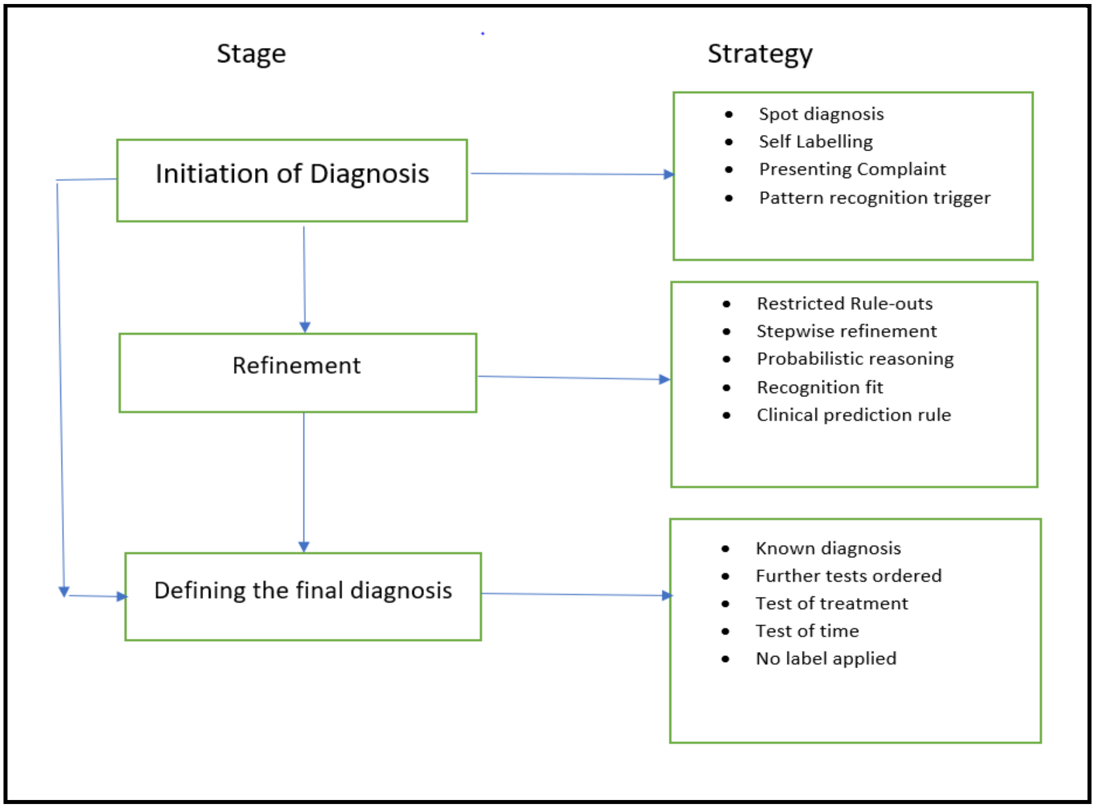

Fig: Different stages in Diagnosis and strategies used by physicians

Initiation of diagnostic hypotheses:

More than one element may be considered during each stage. For instance, some rashes can be diagnosed with a visual spot diagnosis, while for others, spot diagnosis may be part of an overall pattern recognition strategy with probabilistic reasoning—for example, a characteristic chicken pox rash in a child with a fever. In addition, for some diagnoses a high level of certainty at the initiation stage may lead straight to a final known diagnosis, missing out the refinement step—for example, diagnosing simple acne.

The first trigger for a diagnosis usually occurs early in the consultation.

We identified four possible strategies: spot diagnosis, self labelling and pattern recognition trigger.

1. Spot diagnosis: It arises from an unconscious recognition of a particular non-verbal pattern, usually visual (dermatological condition such as acne) or auditory (a barking cough). The

spot diagnosis is almost instantaneous, relies on previous non-verbal experience of the condition, and does not require further history from the patient to trigger the possible diagnosis.

2. Self labelling: The patient may tell you what they perceive to be the diagnosis. This may or may not be correct and is often based on their own or an acquaintance's previous experience of a problem, but it immediately directs subsequent refinement of the diagnosis.
3. Presenting complaint: For instance, "I have abdominal pain" or "I have a headache" was used most often by our General Physicians, and traditional textbooks and teaching recognize this step at the outset of the consultation.
4. Pattern recognition trigger: Elements in the history or examination, or both (sometimes related to the presenting complaint) may trigger the hypothesis. For example, thirst, feeling unwell, and looking unwell in an adolescent trigger the possibility of type 1 diabetes.

Refinement Stage:

Once the initial possible diagnoses are formed, other strategies are used to narrow the possibilities. These strategies are not mutually exclusive. We found that five strategies were used in the refinement stage: restricted rule out process, stepwise refinement, probabilistic reasoning, pattern fit, and clinical prediction rule.

1. Restricted rule outs: This diagnostic strategy depends on learning the most common cause of the presenting problem (the "probability diagnosis") and a shortlist of serious diagnoses which must be ruled out. For instance, in headache the common causes are tension-type headache and migraine, but malignant hypertension, temporal arteritis, and subarachnoid hemorrhage must routinely be ruled out, even if these diagnoses have not been triggered by the presentation. This strategy is aimed at preventing errors in clinical practice.

2. Stepwise refinement: It is based on either the anatomical location of the problem or the putative underlying pathological process. An example of refinement based on the underlying disease is deciding whether conjunctivitis is allergic or infectious.
3. Probabilistic reasoning: It is the specific but probably imperfect use of symptoms, signs, or diagnostic tests to rule in or rule out a diagnosis. Probabilistic reasoning requires knowing the degree to which a positive or negative result of a test adjusts the probability of a given disease. Example- use of electrocardiograms in the assessment of chest pain.
4. Pattern recognition fit: symptoms and signs are compared to previous patterns or cases, and a disease is recognized when the actual pattern fits. This is the refinement strategy most commonly used. Its use relies on memory of known patterns, but no specific rule is used. Some conditions may have various patterns.
5. Clinical prediction rule: It is a formal version of pattern recognition based on a well-defined and widely validated series of similar cases. For instance, HAD score for depression.

Strategies in the Final Diagnosis:

Less than 50% of cases resulted in the certainty of a "known diagnosis" without further testing. The other strategies in the final stage of diagnosis, including ordering further tests, test of treatment, and test of time.

1. Known diagnosis: It is a sufficient level of certainty of the diagnosis to start appropriate treatment or to rule out serious disease without further testing.
2. Ordering further tests: a standard test can sometimes be used to rule in or rule out the disease—for example, midstream urine in urinary tract infection. In addition, further tests were used in response to red flags (Red flags are specific symptoms or signs that may be

volunteered by the patient) and when the diagnoses did not fit any obvious pattern of disease.

3. Test of treatment—when the diagnosis is uncertain, the response to treatment is often used to refute or confirm it. Examples included the use of inhalers in nocturnal cough.
4. Test of time—the course of the disease is used to predict when a person should be better or worse; a ‘wait and see’ strategy allows the diagnosis to become more obvious. For example, in a patient with abdominal pain, diarrhea, and no red flags, and who is diagnosed as having viral gastroenteritis, most Physician’s would wait one or two weeks before considering other disease or testing.
5. No label applied—where no diagnostic label could be assigned to the patient, presentations were often vague and didn’t fit a recognizable pattern. Various strategies can be used recalling the patient for further review or using an exploratory further investigation.

Materials and Methods

Startup companies and their projects:

Healint:

Healint leverages innovative techniques in software, data science and user experience design to empower people to manage their chronic conditions and diseases. Using deep analytics and machine learning, Healint generates unprecedented insights utilized by physicians and pharma companies to help patients on their journey

Healint’s first global program - the Migraine Buddy platform and its apps - helps a thriving community of users manage and track their migraines. They work on collecting, storing, processing, and analyzing the data they receive every week. This also involves helping build machine learning algorithms by preparing and processing training and testing datasets.

HBI Solutions:

HBI solutions deliver actionable information that helps healthcare organizations identify population, quality, and cost risks to improve patient health and lower costs. They use real-time clinical, billing and claims information with built-in natural language processing (NLP) to capture additional clinical, social and behavioral information found in unstructured and non-discrete data types like discharge summaries, histories and physicals.

iCarbonx:

iCarbonX released Meum, a digital health management platform. The company name, “iCarbonX,” symbolizes the use of the internet and artificial intelligence to improve life, of which a central element is carbon. The “i” and “X” indicate the company’s plans to combine the Internet and artificial intelligence to create something new. iCarbonX ranked one of Fast Company’s 2017 Top 10 most innovative in China.

Lumiata:

Lumiata’s analytics offering works with health plans and providers to identify at-risk patients and manage costs. According to the company, its models have so far been used to predict risk or chronic condition onset for more than 20 million patient lives. Lumiata mainly develops automated pipelines for training and deploying machine learning models and building high performance systems to understand complex patient data. They extract and expose external public data sources to enhance medical machine learning models.

Zephyr Health:

Zephyr Health turns data into information for Fortune 100 human health companies. Zephyr Health facilitates efficient and effective data mining/gathering, curation, aggregation, organization, synthesis and visualization to create “intelligent data”; this process involves bringing together disparate data sources and perspectives and applying the right types of “signal-to-noise” filters to drive insight generation that leads to effective action. Zephyr Health’s platform and products are the engine behind more effective decision making, differentiated competitive advantage and optimized customer engagement for life sciences clients. Zephyr Health is developing and leveraging cutting edge technology in a number of areas including database architecture, data mining and modeling, data integration, applied analytics, visualization and natural language processing, to name a few, to develop products and tools that address the unique needs of customers.

Job sites used to determine salary ranges:

Referred From www.linkedin.com

Data Analyst: $120K to 145K

Data Scientist intern: $90K to 110K

Data Scientist: $145K to 160K

Senior Data Scientist: $221K to 260K

Referred From www.glassdoor.com

Data Analyst: $107K to 130K

Data Scientist Intern: $127K to 145K

Quantitative Analyst: $159K to 190K

Senior Data Scientist: $190K to 240K

Referred From www.simplyhired.com

Data Scientist intern: $65K to $88K

Data Scientist: $160K to $210K

Principal Data Scientist: $140K to $190K

Data Analyst: $87K to $120K

Referred From www.ziprecruiter.com

Data Scientist intern: $45K to $62K

Data Scientist: $107K to $160K

HealthCare Data Scientist: $99K to $110K

Data Analyst: $88K to $100K

Companies and their Requirements (Referred From www.paysa.com)

Name of the Company	Position	Skill-Set Required.
Virginia Hospital and Healthcare Association	Data Analyst	• Knowledge on Data Mining and Data Extraction • Tableau and other data visualization tools • Programming languages (C++ or Java or Python)
CVS pharmacy	Data Scientist	• R or Python to manipulate large data sets • Data management in an Hadoop environment, including use of Hive • Specialization in mathematical analysis methods, machine learning, statistical analyses, and predictive modeling

Splunk	Data Scientist	• Algorithms and Data Structures in Python • Fundamentals of AI • Strong Communication Skills
Nuna Health	Data Scientist	• Data Wrangling with MongoDB • Algorithms and Data Structures in Python • Basics of Deep Learning

HealthCare Data Scientist Skill Set:

General Skills

Skill	Use Case Examples
Dimensionality reduction	• Data preprocessing
Supervised Machine Learning	• Classification of patients • Forecasting of numeric values (e.g. length of treatment number of staff needed on a shift)
Time Series Analysis	• Hospital Management • Forecasting the growth in a certain healthcare sector
Natural language processing	• Disease prediction • Forecast of disease complications • Defining specific patient groups for research • Improving quality and integrity of medical records

	• Claims processing. Feedback analysis

Healthcare-specific skills

Skill	Use Case Examples
Medical coding classification systems	• Analysis of patient records • Billing operations and reimbursement • Mortality and morbidity statistics
Healthcare Databases	• Public health forecasting • Hospital Management • Drug research and development.

Quantitative Methods

Machine learning in HealthCare:

From our findings, here is a list of the most common techniques and ML algorithms employed in Healthcare Analytics:

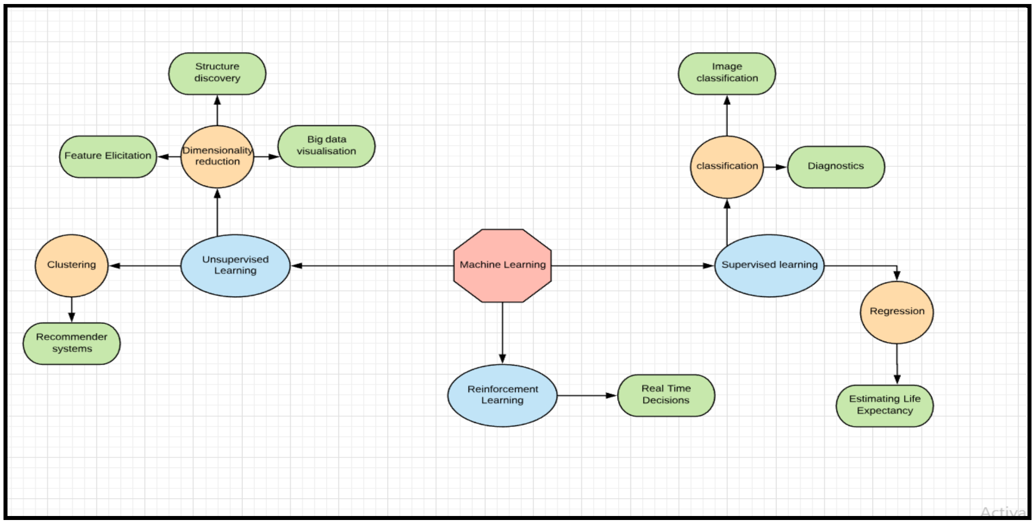

Fig: Overview of Machine learning techniques in Health care

Some popular examples of supervised machine learning algorithms used in Clinical diagnosis are:

- Logistic regression for regression problems.
- Random forest for classification and regression problems.
- Support vector machines for classification problems.

Logistic Regression:

We looked into three healthcare diagnostics classification datasets from Kaggle: Wisconsin diagnostic Breast Cancer dataset, Pima Indians Diabetes Dataset and Indian Liver Patient Dataset. And explored different Kaggle kernels to find out different methods and algorithms used to classify the Dataset. In our exploration, we encountered various methods as Random Forest , Support Vector Machine, Logistic Regression and, to name a few . In this paper we are discussing in detail about the Logistic Regression.

The combination of medical judgement and an algorithmic diagnostic tool based on extensive medical records is the future of medical diagnosis and treatment. Logistic regression has one big thing going for it – a lot of logistic regressions have been performed to identify risk factors for various diseases or for mortality from a particular ailment.

Logistic regression is the appropriate regression analysis to conduct when the dependent variable is dichotomous (binary). Like all regression analyses, the logistic regression is a predictive analysis. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables.

The problems of picking the correct explanatory variables for a logistic regression and model validation are linked. We have to try various specifications (sets of explanatory variables) and utilize a raft of diagnostics to evaluate the different models. Cross-validation, utilized in the breast cancer research mentioned above, is probably better than in-sample tests.

In logistic regression we have a kind of background relationship which relates an odds-ratio to a linear predictive relationship, as in,

$\ln(p/(1-p)) = \beta_0+\beta_1x_1+\beta_2x_2$

This **odds ratio** is really primary and from the logarithm of the odds ratio we can derive the underlying probability p. This probability p, in turn, governs the mix of values of an indicator variable Z which can be either zero or 1, in the standard case (there being a generalization to multiple discrete categories, too).

The parameters β_i in a logistic regression are estimated by means of **maximum likelihood (ML)**. Among other things, this can mean the optimal estimates of the beta parameters – the parameter values which maximize the likelihood function – must be estimated by numerical analysis, there being no closed form solutions for the optimal values of β0, β1, and β2 (Jones, n.d.).

In addition, interpretation of the results is intricate, there being no real consensus on the best metrics to test or validate models. SAS and SPSS as well as software packages with smaller market shares of the predictive analytics space, offer algorithms, whereby we can plug in data and pull out parameter estimates, along with suggested metrics for statistical significance and goodness of fit.

Random Forest:

Random forests start with the idea of decision tree. This is a way of deciding based on a flow chart. The random forest is a way of deciding based on votes from many decision trees which are created by using a subset of the attributes. Random Forest is a type of Ensemble method, which combines several machine learning techniques in to one predictive model to decrease variance, bias or improve predictions. Rather than splitting a tree node using all variables, RF selects at each node of each tree, a random subset of variables, and only those variables are used as candidates to find the best split for the node.

Support Vector Machine (SVM):

A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. Support Vector Machine is a Machine Learning tool used for classification that is based on Supervised Learning which classifies points to one of two disjoint half-spaces. It uses nonlinear mapping to convert the original data into higher dimension. Its objective is to construct a function which will correctly predict the class to which the new point belongs, and the old points belong.

Co-Forest:

The semi-supervised learning has been widely applied in many fields such as medical diagnosis, pattern recognition. The semi supervised learning methods are used to employ unlabeled data in addition to labelled data for better classification of large data sets, where only a small number of labelled examples is available. Co-Forest is most commonly used semi-supervised method, which is an implementation of random forest in semi-supervised model.

Data Mining in HealthCare:

Data mining is the process of pattern discovery and extraction where huge amount of data is involved. Data mining is gaining popularity in different research arenas due to its infinite applications and methodologies to mine the information needed. Healthcare industry today produces huge amounts of multi-faceted data from hospitals, resources, disease diagnosis, electronic patient records, etc. The large amount of data is crucial to be processed and scrutinized for knowledge extraction that empowers support for understanding the prevailing circumstances in healthcare industry. Data mining processes include framing a hypothesis, gathering data, performing pre-processing, estimating the model, and understanding the model and draw the conclusions.

Table: Summary of medical data mining techniques

Disease	**Techniques used**
Conventional Pathology Data	Extracting patterns & detecting trends using Neural Networks.
Coronary heart disease	Prediction models using Decision Tree Algorithms such as ID3, C4.5, C5, and CART.
Lymphoma Disease and Lung Cancer	Distinguish disease subtypes using Ensemble approach.
Psychiatric Diseases	Predicate the probability of a psychiatric patient on the basis detected symptoms using BBN Bayesian networks
Fre quent Disease	Identify frequency of diseases in particular geographical area using Apriori algorithm.
Liver diseases	Classification using Bayesian Ying Yang
Skin Disease	Categorization of skin disease using integrated decision tree model with neural network classification methods
Diabetes	Classification of Medical Data using Genetic Algorithm
Functional Magnetic Resonance Imaging (fMRI)	Integration of Clustering and Classification of biomedical databases
Chest Disease	Constructed a model using Artificial Neural Network
Diabetes, Cancer	Classification of Disease using k-Nearest Neighbor
Coronary Heart Disease	Improving classification accuracy using Naive Bayesian
Chronic Disease	Prediction of Diseases Using Apriori Algorithm
Diabetes	Disease classification using Support Vector Machine
Breast Cancer	Accurate Classification of medical data using K-means, Self-Organizing Map (SOM) and Naïve Bayes
Cardio Vascular Diseases	Diagnose Cardio Vascular Disease using Classification algorithm

Parkinson Disease	Familiarized an adaptive Fuzzy K-NN approach for diagnosing the disease

Big data in Healthcare:

What Is Big Data in Healthcare:

Big Data has changed the way we manage, analyze and leverage data in any industry. Healthcare analytics have the potential to reduce costs of treatment, predict outbreaks of epidemics, avoid preventable diseases and improve the quality of life in general. Health professionals, just like business entrepreneurs, can collect massive amounts of data and look for best strategies to use these numbers. Big data refers to the vast quantities of information created by the digitization of everything, that gets consolidated and analyzed by specific technologies.

Why We Need Big Data Analytics in Healthcare:

Physician decisions are becoming more and more evidence-based, meaning that they rely on large swathes of research and clinical data as opposed to solely their schooling and professional opinion. As in many other industries, data gathering, and management is getting bigger, and professionals need help in the matter. This new treatment attitude means there is a greater demand for big data analytics in healthcare facilities than ever before.

Big Data Applications in Healthcare:

1. Electronic Health Records (EHRs):

   This is the widespread application of Big Data. This includes the patient’s demographics, medical history, allergies, laboratory test results etc. thus every patient has his/her own digital record. Every record is comprised of one modifiable file, which means that doctors can implement changes over time with no paperwork and no danger of data replication.

2. Real-Time Alerting:

   Other examples of big data analytics in healthcare share one crucial functionality – real-time alerting. In hospitals, Clinical Decision Support (CDS) software analyzes medical data on the spot, providing health practitioners with advice as they make prescriptive decisions. An application of real-time Alerting: In Asthma polis patients, they started to use inhalers with GPS-enabled trackers to identify asthma trends both on an individual level and looking at larger populations. This data is being used in conjunction with data from the CDC to develop better treatment plans for asthmatics.

3. Enhancing Patient Engagement:

   Many consumers already have an interest in smart devices that record every step they take, their heart rates, sleeping habits, etc., on a permanent basis. All this vital information can be coupled with other trackable data to identify potential health risks lurking. A chronic insomnia and an elevated heart rate can signal a risk for future heart disease for instance. Patients are directly involved in the monitoring of their own health, and incentives from health insurances can push them to lead a healthy lifestyle (e.g.: giving money back to people using smart watches). Patients suffering from asthma or blood pressure could benefit from it and become a bit more independent and reduce unnecessary visits to the doctor.

4. Big Data Might Just Cure Cancer:

   Another interesting example of the use of big data in healthcare is the Cancer Moonshot program. Medical researchers can use large amounts of data on treatment plans and recovery rates of cancer patients to find trends and treatments that have the highest rates of success in the real world. For example, researchers can examine tumor samples in biobanks that are linked up with patient treatment records. Using this data, researchers can see things like how certain mutations and cancer proteins interact with different treatments and find trends that will lead to better patient outcomes.

5. Patients Predictions for An Improved Staffing:

   Doctors, Nurses and Hospital administration staff use data from a variety of sources to come up with daily and hourly predictions of how many patients are expected to be at each hospital, to forecast visit and admission rates for the next 15 days. Extra staff can be drafted in when high numbers of visitors are expected, leading to reduced waiting times for patients and better quality of care.

6. Predictive Analytics in Healthcare:

   Predictive analytics is recognized as one of the biggest business intelligence trend two years in a row. The goal of healthcare business intelligence is to help doctors make data-driven decisions within seconds and improve patients' treatment. This is particularly useful in case of patients with complex medical histories, suffering from multiple conditions. New tools would also be able to predict, for example, who is at risk of diabetes, and thereby be advised to make use of additional screenings or weight management

7. Integrating Big Data with Medical Imaging:

   Medical imaging provider Carestream explains how big data analytics for healthcare could change the way images are read: algorithms developed analyzing hundreds of thousands of images could identify specific patterns in the pixels and convert it into a number to help the physician with the diagnosis. They even go further, saying that it could be possible that radiologists will no longer need to look at the images, but instead analyze the outcomes of the algorithms that will inevitably study and remember more images than they could in a lifetime. This would undoubtedly impact the role of radiologists, their education and required skillset.

Conclusion

From predicting treatment outcomes, to curing cancer and making patient care more effective, data science healthcare has proven to be an invaluable contribution to the future of the industry. Data Science has the power to transform the way healthcare providers use sophisticated technologies to gain insight from their clinical and other data repositories and make informed decisions. From our research, we identified that the boost in health innovation is driven by the three main factors:

- Advances in technology
- Growth of digital consumerism
- The need to fight increasing costs

While data science provides tools and methods to extract real value from unstructured patient information, it eventually contributes to making healthcare more efficient, accessible and personalized. The number of healthcare institutions making data-driven decisions increases slowly but steadily. In 2015, only 15 percent of hospitals employed data science and predictive analytics to prevent hospital readmissions. One year after, 31 percent of institutions said they have been doing so for more than a year.

Apart from the current advances in technology in healthcare, it was noticed there are few areas where a lot of work is still required. Healthcare data is largely fragmented and siloed. This is in part due to late adoption of EHR technology across the industry. Common data architecture is a stepping stone in the healthcare sector, as data sharing has become complicated. Also, the lack of common standards of interoperability as well as the high amount of effort it takes for IT personnel to design and build safe and effective data integration between various sources is a culprit. Strong data governance efforts have been inadequate in this sector. All these issues must be addressed. By overcoming these, Healthcare is poised to bring in great benefits for patients, providers and payers. Under the health care umbrella there are numerous opportunities for Data scientists.

Here are few important points from our Discussions:

- Startups vs MNC's

  Data science and machine learning is a very hot area for AI tech start-ups. Clinical informatics is also high in demand and is critical to building analytics in the healthcare space. Health care startups are competing with consumer companies like Google, Amazon, and Uber and the talent supply is not keeping up with demand. Typically, these companies have higher budgets, appealing brand names compared to the startups, especially Health care startups are still behind in technology compared to consumer business. People with a humanistic underpinning are drawn to the mission and, purpose and are excited about joining a health care startup team culture that is passionate about improving healthcare.
- Amount of venture capital money flowing into health tech change the dynamics significantly:

  The increased capital has helped, but with additional capital comes increased pressure to grow quickly. It's a balancing act. To retain a lead in the market and consistently innovate, you need great talent and an infrastructure that scales over time and doesn't remain linear. Market penetration and relevance are the keys to allowing the top line to increase while giving the time to scale and manage expenses. This will allow startups to expand their development and marketing efforts into many new adjacencies which supports accelerated growth while expanding markets.

- Most sought after roles and skillsets among Health Tech start up's and Health care professional's retention rate:

  Depending on location, it is difficult to find certain tech and healthcare talent. Average tech retention rate of most of the start up's is more than twice that of Apple or Google.

  Data scientists with PhDs are highly in demand and their most common fields of study are Mathematics and Statistics (25%), followed by Computer Science (20%), Natural Sciences such as Physics (20%), and Engineering (18%). The next preferred are the ones with master's degree in the related domains. From the research, we could conclude that we have to work on our python skills, ML techniques, knowledge on Hadoop platform and unstructured data apart from the intellectual curiosity, business acumen and communication skills the data scientists are built in with.

References

Baum, S. (n.d.). *medicity news*. Retrieved from https://medcitynews.com/2018/06/the-race-to-find-data-scientists/?rf=1

C, H. (n.d.). Retrieved from bmj: https://www.bmj.com/bmj/section-pdf/186199?path=/bmj/338/7701/Practice.full.pdf

Chen, X. (n.d.). *ncbi*. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3387489/

Jones, C. (n.d.). *businessforecastblog*. Retrieved from http://businessforecastblog.com/medicalhealth-predictive-analytics-logistic-regression/

Lii, M. (n.d.). Retrieved from ieee: https://ieeexplore.ieee.org/document/4342802

Patel, S. (n.d.). Retrieved from http://aircconline.com/ijist/V6N2/6216ijist06.pdf

Torry, T. (n.d.). *verywellhealth*. Retrieved from https://www.verywellhealth.com/primary-secondary-tertiary-and-quaternary-care-2615354