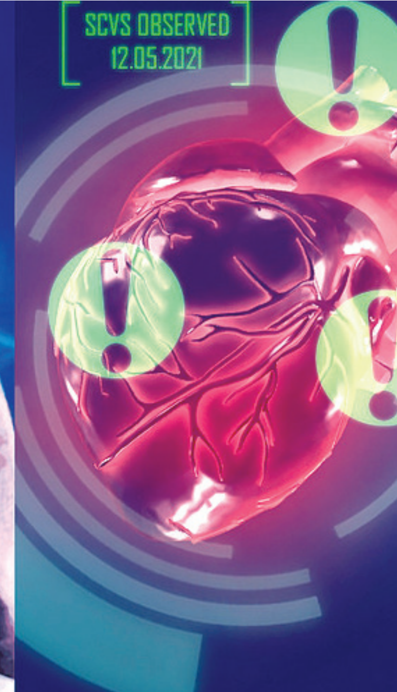
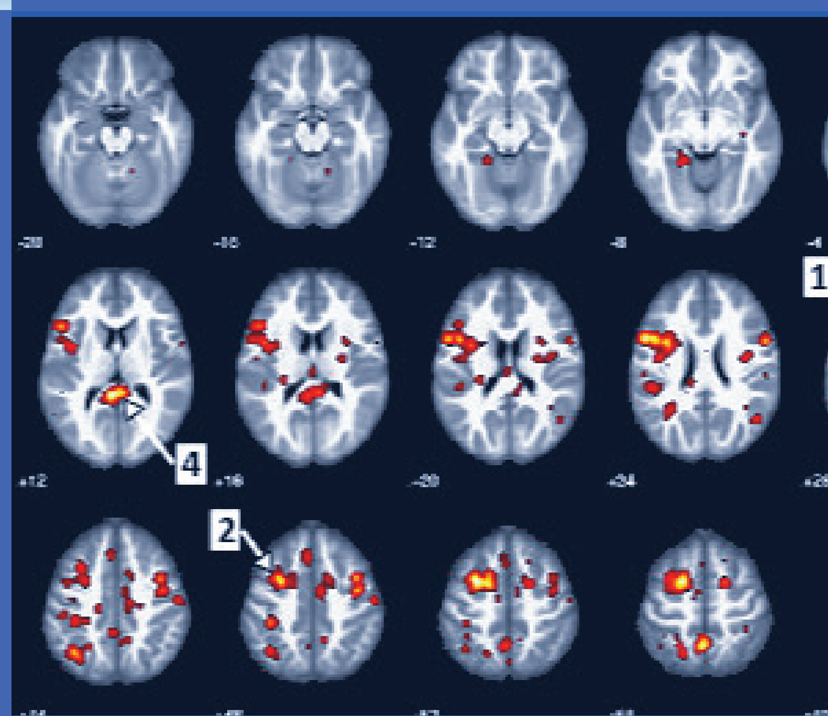
ActiveState®



MACHINE LEARNING IN HEALTHCARE



RESEARCH TO REALITY

"Big data and machine learning in pharma and medicine could generate a value of up to $100B annually, based on better decision-making, optimized innovation, improved efficiency of research/clinical trials, and new tool creation for physicians, consumers, insurers, and regulators."

McKinsey

CHAPTER 1

UNDERSTANDING ML ROLE IN HEALTHCARE

While Artificial Intelligence (AI), which is defined as the ability for a machine to make cognitive decisions without human input has garnered a lot of attention recently, it's Machine Learning (ML) that is actually delivering on the promise of what AI may someday achieve. ML allows computers to progressively improve their performance by learning how to do some automated task over time, typically a task that would previously have required human intervention.

But despite the great strides that researchers have made in the field of ML, it doesn't mean we'll all soon be replaced by machines. On the contrary, ML is a just a tool, and like any tool it can only be as effective as the user. This guide is designed to help you better understand ML in order to be able to wield it more effectively in your healthcare business.

ML has been a component of healthcare research since the 1970's when it was first applied to tailoring antibiotic dosages for patients with infections. But with the increased volume of Electronic Health Records (EHR) and the explosion in genetic sequencing data, healthcare's interest in ML is now at an all-time high. In fact, no other industry exhibits ML investments on the same scale as healthcare from 2015 through 2017. **(See Figure 1.)**

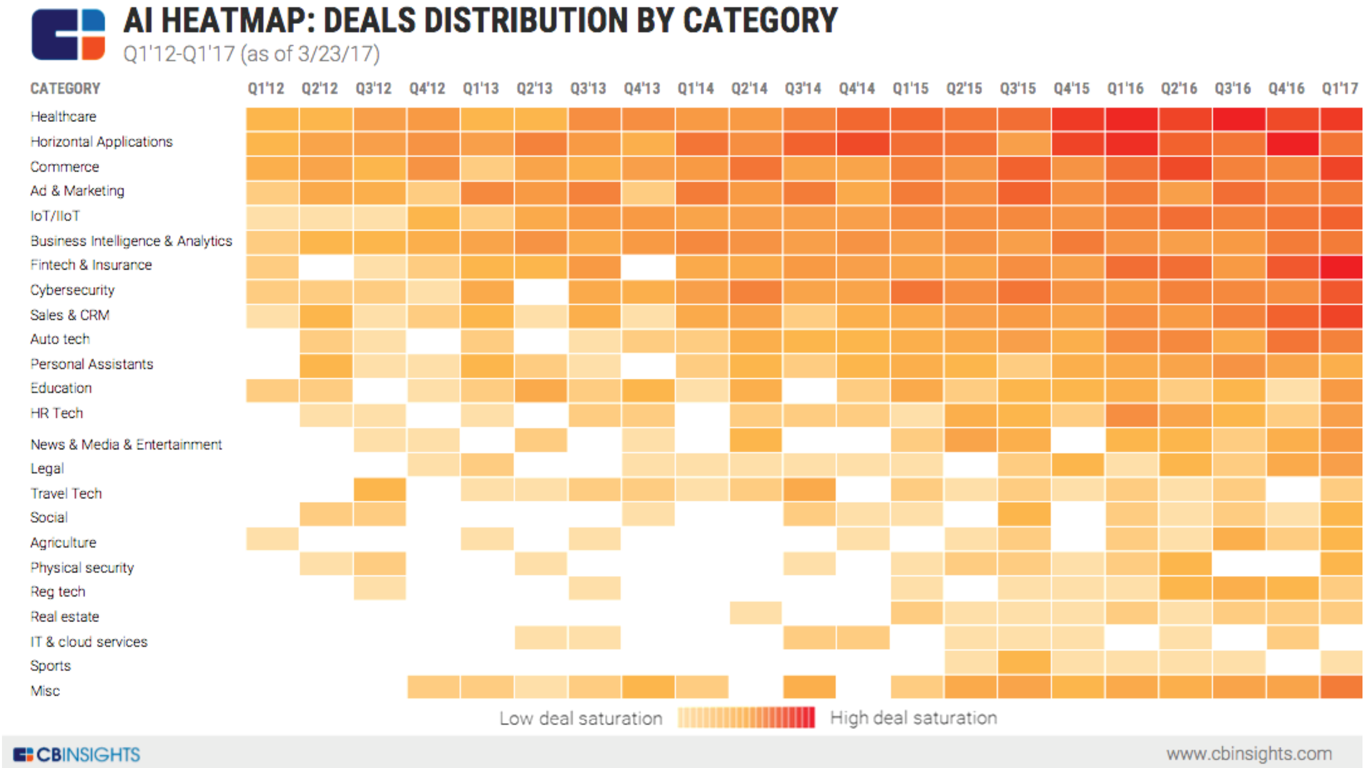


FIGURE 1: INVESTMENT IN ML BY INDUSTRY

McKinsey sums up this growing interest succinctly: "ML in pharma and medicine could generate a value of up to $100B annually, based on better decision-making, optimized innovation, improved efficiency of research/clinical trials, and new tool creation for physicians, consumers, insurers, and regulators." [1]

Both ML and healthcare are enormous fields unto themselves. This guide focuses on ML as it applies to the following aspects of healthcare:

- ML in Public Health Data Analysis
- ML for Healthcare Management
- ML in Genetics
- ML in Diagnostics

[1] https://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/how-big-data-can-revolutionize-pharmaceutical-r-and-d

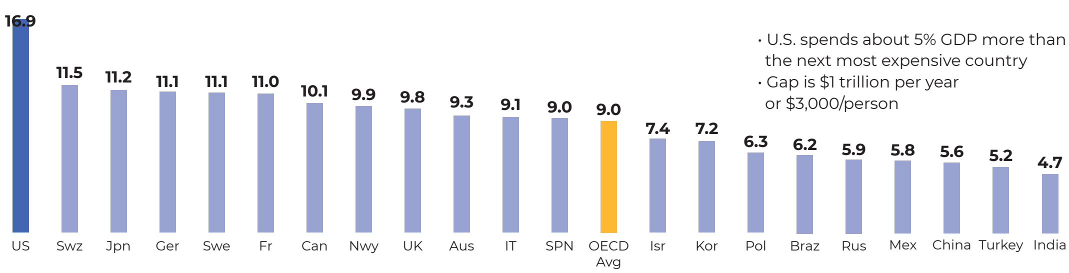
STATE OF THE INDUSTRY

Today the healthcare industry – from public health to hospital administration to research and diagnostics – is growing at a phenomenal rate as baby boomers age and medical technologies continue to provide significant advancements. ML has a potential role to play in all these areas, while helping to address one of the industry's most pressing concerns: rising costs.

According to the Organization for Economic Co-operation and Development (OECD), healthcare spending averages 9% of GDP for 34 of the 35 OECD countries. In the the US, the 35th country, that number is closer to 17% of GDP. On average, healthcare costs for the 34 OECD countries are growing at 4.9% per year, which exceeds the U.S. rate. By comparison, in China, a non-OECD country, healthcare spending is growing at a rate of 16%.

Skyrocketing costs are forcing the industry to do more with less, which is why digital transformation through technological innovation is key. ML is proving to be one of the best ways to optimize operations and maximize cost efficiency.

2015 HEALTHCARE EXPENSES (% GDP)



http://stats.oecd.org/Index.aspx?DataSetCode=SHA

FIGURE 2: OECD HEALTH STATISTICS 2015

ML in Cancer Care

University of Texas MD Anderson Cancer Center (MDACC) has been training IBM's Watson to improve the quality of care for cancer patients with no access to cancer specialists. The result is a virtual expert called MD Anderson Oncology Expert Advisor (or OEA) that provides both therapy decisions and recommendations to maximize treatment benefits. In other words, OEA is designed to share both the clinical evidence as well as the "art" of cancer care.[2]

[2] http://www.tomdavenport.com/wp-content/uploads/Lessons-from-the-Cognitive-Front-Lines-Early-Adopters-of-IBMs-Watson.pdf

INTRODUCING ML

Analytics has traditionally been used to help discover insights in data. But analytics has proven ineffective when it comes to multi-dimensional and/or noisy data. This is where ML comes in, helping analysts identify actionable data patterns in complex datasets. The two most popular ways to apply ML are using supervised and unsupervised learning.

Supervised learning requires a large source of classified information, which typically means that human experts have already examined and labeled the data. For example, X-rays that have been classified into various categories of fractures versus breaks versus dislocations, etc. Using ML data scientists can then apply algorithms to the data in order to learn to classify it just as well as, or perhaps even better than, subject matter experts. In diagnostics this means finding patterns in clinical data that correlate with expert diagnoses. Or it may be used to identify patients who are at risk of particular post-treatment outcomes, and so on.

Unsupervised learning uses algorithms to find patterns in data that has NOT been previously classified. For example, unsupervised learning is the primary method used to identify the mechanisms for multi-factorial diseases. Unsupervised ML projects apply clustering algorithms in order to find ways to partition the data into some number of more-or-less isolated groups (i.e. clusters).

ML in Radiology and Radiotherapy

Google's DeepMind Health is working with University College London Hospital (UCLH) to develop machine learning algorithms capable of detecting differences in healthy and cancerous tissues to help improve radiation treatments. ML speeds up the segmentation process (ensuring that no healthy structures are damaged) and increases accuracy in radiotherapy planning. [3]

[3] https://www.techemergence.com/machine-learning-in-pharma-medicine/

"Today, some surgical procedures already incorporate smart machines. Robots routinely carry out the crucial steps in some procedures including orthopaedic knee replacements, Lasik eye surgery, and hair transplants."

The Human OS

CHAPTER 2

DATA ANALYSIS IN PUBLIC HEALTH

No matter whether you choose to work with a supervised or unsupervised ML method, you will require data. Lots of data. If you haven't already done so, first on your radar should be separating your data from your applications, and placing that data in a distributed, shareable repository. This has two benefits: 1) geographically disparate healthcare systems can leverage a single source of truth, and 2) it provides ML teams with the data they require to test and train their algorithms against.

If you don't have suitable data to work with today, you can still get started with ML by leveraging both free and pay-for sources of public health data, such as:

- **Big Cities Health Inventory Data** – provides health data from 26 cities for 34 health indicators across six demographic indicators.
- **Healthcare Cost and Utilization Project (HCUP)** – identify, track, and analyze trends in health care utilization, access, charges, quality, and outcomes at the US national and state levels.
- **HealthData.gov** – constantly updated health-related data sets from across the US. Provides the ability to search and export data.
- **Surveillance, Epidemiology & End Results (SEER)-Medicare Health Outcomes Survey (MHOS)** – cancer patient diagnoses, clinical information and outcomes database.
- **The Human Mortality Database (HMD)** – detailed mortality and population data for over 35 countries.
- **Child Health and Developmental Studies** – data investigating how health and disease are passed on between generations.
- **Medicare Provider Utilization and Payment Data** – information about services and procedures provided to US Medicare beneficiaries.

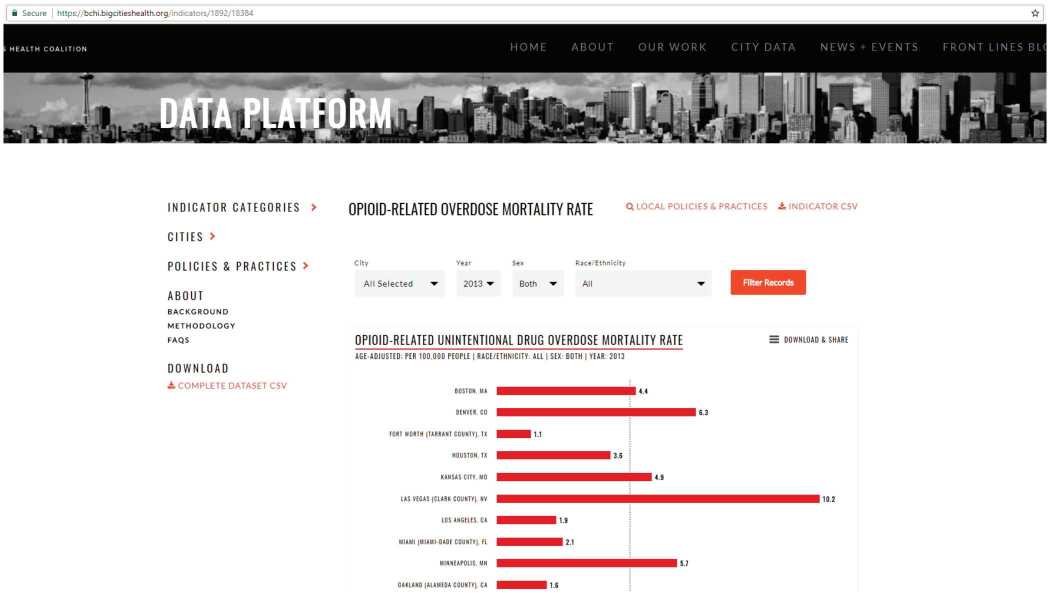


FIGURE 3: BIG CITIES HEALTH DATA PLATFORM

Data analysis has played a big role in public health for over 150 years. John Snow's careful analysis of the 1854 cholera outbreak in London was the first analysis of public health data. Although his analysis was followed by several decades of vehement political opposition by miasma theorists, John's work ultimately led to a deeper understanding of the spread of waterborne disease.

Today, ML is being applied to help find insights in vastly larger and more complex datasets whose high degree of dimensionality defies the simple analysis that a John Snow or traditional analytics tool can bring to bear. While not always successful, ML has been applied to everything from predicting flu outbreaks[4] to understanding "global disease burden"[5] and using the insights to better focus resources on ways that do the most good.

ML and Medical Imaging

Microsoft's InnerEye project builds on years of research in computer vision and ML to perform automatic, quantitative analysis of three-dimensional radiological images. The resulting images are so precise they can be effectively turned into measuring devices that allow for better targeted radiomics measurements when planning radioactive therapy.[6]

[4] https://www.wired.com/2015/10/can-learn-epic-failure-google-flu-trends/
[5] https://sinews.siam.org/Details-Page/machine-learnings-impact-on-global-public-health
[6] https://www.microsoft.com/en-us/research/project/medical-image-analysis/

ML OBJECTIVITY

The apparently simple question "What human deaths can be practically prevented?" turns out to be extremely complex. Plus, answers can be affected by human biases. This is particularly because public health issues are necessarily political issues. For example, miasma theorists held significant political power at the time of John Snow's analysis. Miasma theorists believed diseases were caused by "bad air" arising from decaying organic matter. It took several decades for the careful, systematic data analysis done by Snow and others to overcome that resistance. Today, drug use, cloning and stem cell research are equally contentious in many jurisdictions.

Whether these biases stem from political, philosophical, social or other cognitive biases, we have a tendency to get attached to our hypotheses, and that can bend our thinking in a variety of ways. ML provides an alternative. With careful thought and independent review of training data, ML can reduce or even eliminate the impact of biases by relying on algorithmic – rather than cognitive – decision making. So, as well as dealing with the complex, high dimensionality datasets inherent in public health studies, ML can also deliver insights with a higher degree of objectivity.

ML and Surgery

Smart surgical robots won't replace humans in the operating room anytime soon, but the Smart Tissue Autonomous Robot (STAR) has been shown to outperform human surgeons who were given the same task: stitching up a pig's small intestines. STAR solved the challenge using a specialized camera and vision system, creating its own plan for the suturing job, and adjusting that plan as tissues moved during the operation.[7]

---

[7] https://spectrum.ieee.org/the-human-os/robotics/medical-robots/autonomous-robot-surgeon-bests-human-surgeons-in-world-first

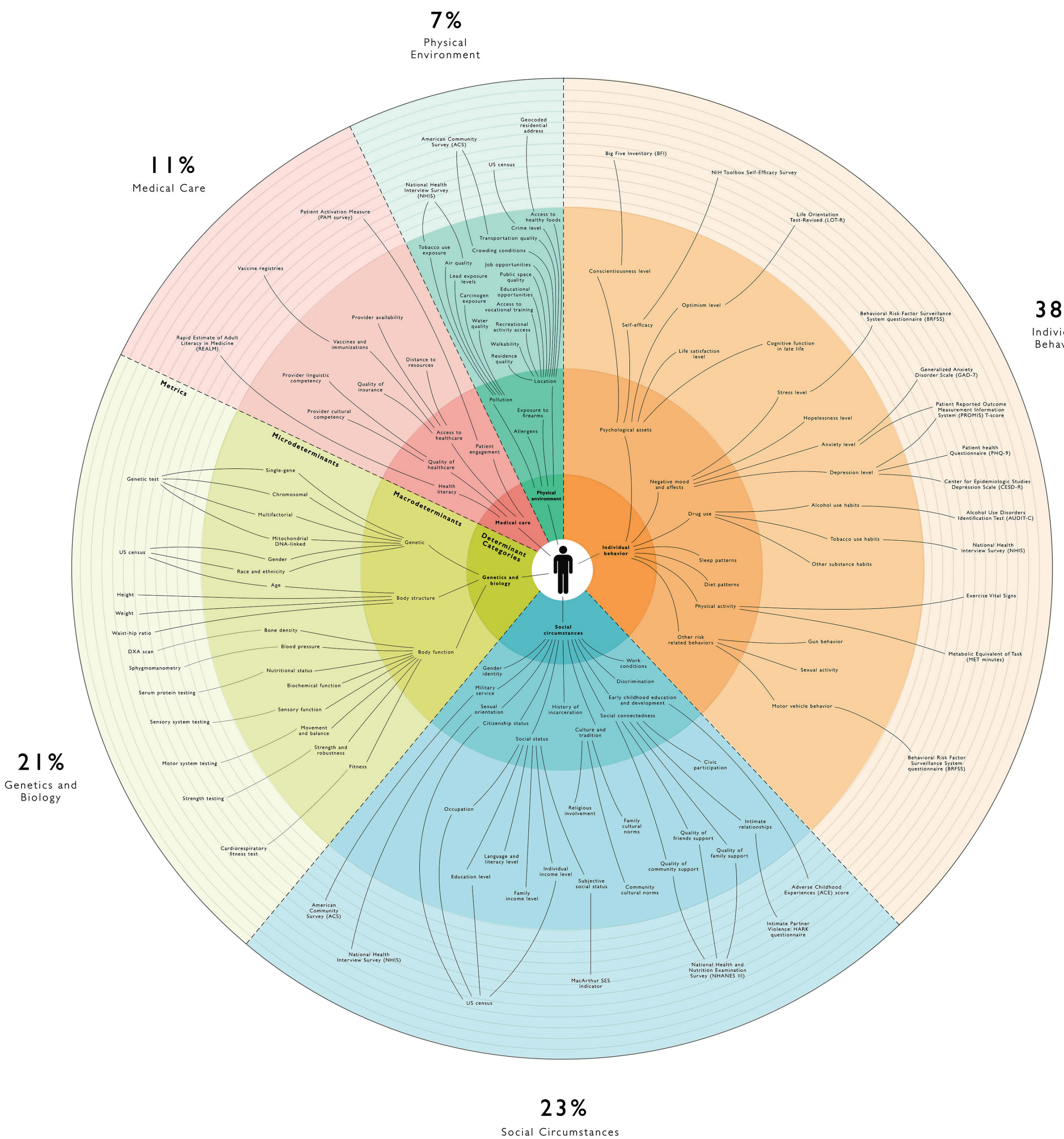


FIGURE 4: SOCIAL DETERMINANTS OF HEALTH

"Ever since healthcare was something humans did, the patient has had to hold up his hand and the system would respond. The idea [with ML] is that if you have rich enough data you can instead predict who may need help and do outreach and move care upstream."

Don Mordecai, M.D., Kaiser Permanente National Leader for Mental Health and Wellness

CHAPTER 3

MANAGE HOSPITALS & CARE FOR PATIENTS

Hospitals and healthcare clinics deliver extremely complex human services under trying conditions. As a result, there is no lack of challenges with patient treatment, as well as general hospital management, all of which can have a direct impact on costs, outcomes and customer satisfaction.

For example, because hospitals and clinics are strongly resource-constrained, cost control is critical to sustainability and long-term effectiveness. And ensuring medical staff, treatment, and diagnostic facilities are scheduled efficiently is a large-scale optimization problem of high dimensionality.

For patients, one of the key determinants of successful outcomes is adherence to prescribed treatment. But people lie, forcing physicians to develop various methods for identifying patients who are not adhering to their treatment protocol.

In addition, patients that require operations also require the attention of many of the hospital's highest paid individuals. Surgical robots have the potential to both reduce surgeon and operating room booking times, as well as provide better outcomes.

ML can play a role in all these areas, from predictive inventory management to improved triage for emergency departments to patient surgery and care.

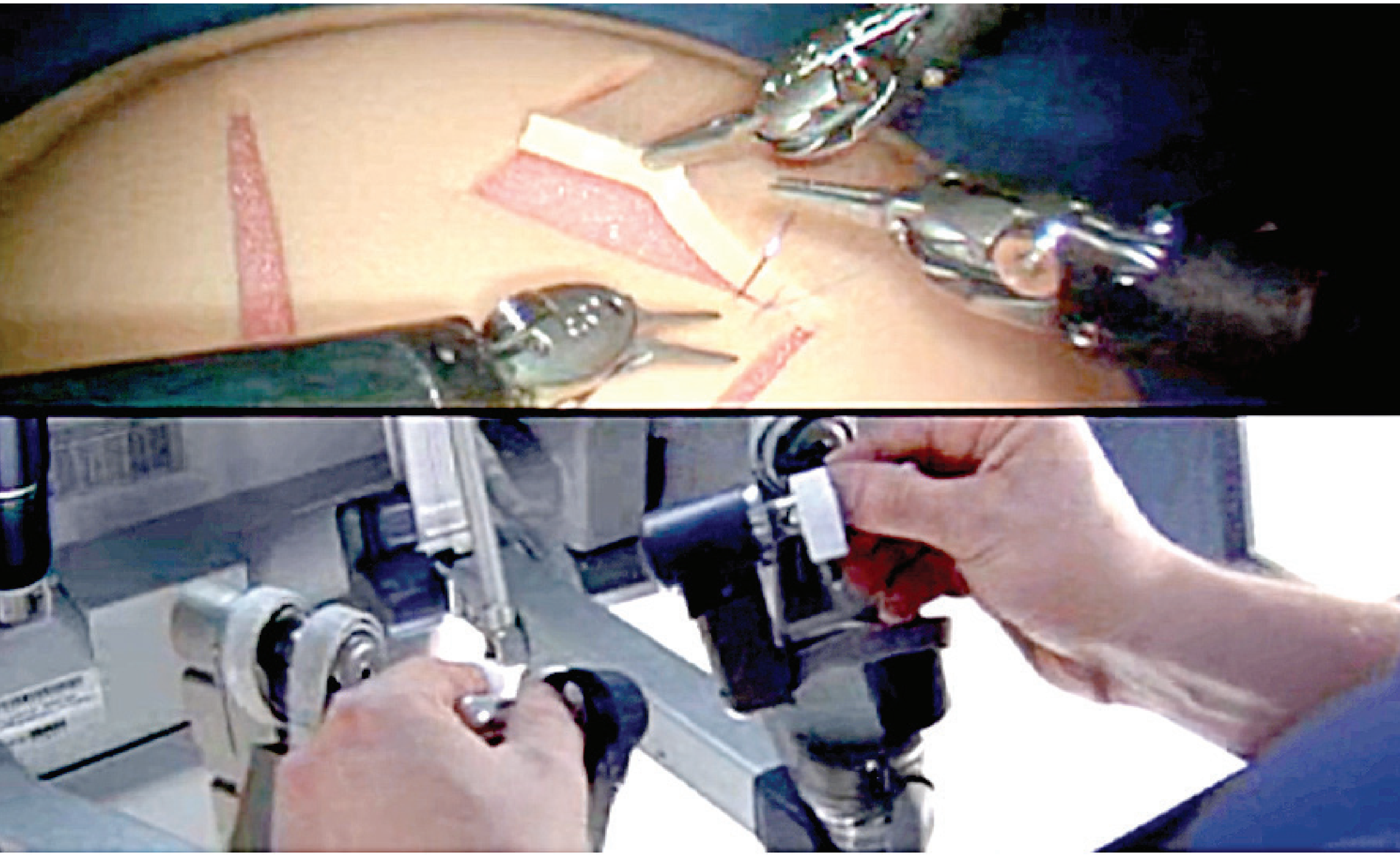


FIGURE 5: ROBOT-ASSISTED SURGERY

ML in Patient Care

ML approaches have proven to be reasonably effective (for specific cases) in identifying non-adherent patients. In the referenced study, people with Parkinson's disease who were non-adherent could be identified with up to 97% accuracy. In this way, ML can identify patients who are unlikely to be compliant with physician advice, enabling interventions. As a result, doctors can not only treat and advise patients, but now also have the means to ensure their advice is followed. [8]

[8] https://www.ncbi.nlm.nih.gov/pubmed/26406881

ML'S ROLE IN COST REDUCTION

Industry analysts estimate that by 2026 the ML health market could potentially save the U.S. healthcare economy $150 billion in annual savings. (See Accenture analysis on the following page.)

Using ML to find patterns and predict outcomes is relatively straightforward. Patient wait times in the ER can be predicted based on staffing levels, time of year, and other factors. There is a great deal of data already available to hospital IT departments that can be fed into ML algorithms and used to train models based on measured outcomes. This can be used for a variety of predictive management tasks.

For example, current drug inventories in hospital pharmacies are known and tracked. Using ML to look at the rise and fall of inventories over time may help achieve greater savings and lower wastage. A hospital IT department could get closer to "just in time" inventory management based on everything from time of year to the type and schedule of surgeries.

Surgical robots that use ML to mesh patient medical records with real-time operating metrics not only enhance a physician's instrument precision but can also provide insights that result in improved techniques. A recent analysis by Accenture also showed that robotic outcomes include a 21% reduction in length of patient stay.

ML-based medical advisors such as virtual nursing assistants can assess a patient's symptoms over the phone and provide advice, saving up to 20% of a nurse's time and also reducing unnecessary hospital visits. In a clinical setting, voice-enabled symptom checkers triage patients and can direct them to the next step.

Other ML-based assistants provide time savings by automating administrative workflow. For example, automated voice-to-text transcription can eliminate the writing of chart notes, test ordering, and even prescription writing. Accenture analysis shows that this kind of ML-based workflow automation can save as much as 17% of doctors' time and up to 51% of nurses' time.

[9] https://www.accenture.com/us-en/insight-artificial-intelligence-healthcare

ML in Nursing

Virtual care company Sense.ly provides in-home nurse avatars that use ML to answer health-related questions based not only on symptoms, but also patient mood and behavior. For those times when more expert care is required, Sense.ly connects you to a doctor in real time via phone, tablet, TV or computer. Sense.ly can also feed wired and wireless medical device data to clinicians, but may soon be able to process that data locally using ML in order to assess risk, triage and coordinate a care plan.

APPLICATION	VALUE*	APPLICATION	VALUE*
Robot-Assisted Surgery**	$40B	Connected Machines	$14B
Virtual Nursing Assistants	$20B	Clinical Trial Participant Identifier	$13B
Administrative Workflow Assistance	$18B	Preliminary Diagnosis	$5B
Fraud Detection	$17B	Automated Image Diagnosis	$3B
Dosage Error Reduction	$16B	Cybersecurity	$2B

TOTAL= ~$150B

Source: Accenture analysis

* "Value" is the estimated potential annual benefits for each application by 2026

** Orthopedic surgery specific

[10] https://hitconsultant.net/2015/06/24/mindmeld-sense-ly-build-artificial-intelligence-health-apps/

"The market for health-related AI is projected to grow at a compound annual growth rate of 40% to $6.6B in 2021 starting from around $600M in 2014."

Accenture Consulting, Artificial Intelligence: Healthcare's New Nervous System

CHAPTER 4

PATTERN MATCHING IN GENE RESEARCH

Looking for patterns of gene expression that correlate with disease state is one of the most basic medical research activities in the past twenty years. However, there are over 30,000 known coding regions in the human genome, which handle the synthesis of over ten times that many proteins. As a result, genomic analysis has become a hotbed for ML-based solutions that are ideally suited to address these kinds of complex, high-dimensionality problems.

There are two basic types of genomic analysis suited to ML: gene sequencing and gene expression.

ML IN GENE SEQUENCING

Sequence data is the static state of the genome: the fixed sequence of base pairs that make up the chromosomes. Small variations in the sequence of specific genes can result in dramatically different expressions in individuals. For example, a single base-pair deletion at any one of positions 35, 167 or 235 in the gene GJB2 is the cause of about half of all hearing loss in humans.

Researchers have been using ML to identify sequence patterns within genetic data sets for decades. But it's only recently that advances in technology have dramatically lowered the cost of genomic sequencing, resulting in a boom in gene pattern research. These patterns are then translated to computer models designed to help predict the probability of an individual developing a specific disease or to aid in the design of therapies/drug candidates.

Gene editing (i.e., using CRISPR technology to alter DNA) also benefits from ML. Gene editing allows for the repair of faulty genes, and may soon lead to the eradication of debilitating and fatal genetic diseases such as Cystic Fibrosis, Parkinson's Disease and Alzheimer's Disease.

Because gene editing is still in its infancy it's hard to predict the effect of any one CRISPR experiment. Most physiology is the result of complex multi-gene effects which increase the dimensionality of the data. Looking at how the effect of naturally occurring mutations and/or man-made edits to one gene interact with variations in another gene is a problem that can only be approached using ML. However, as more data is collected, better ML models can be created, which will lead to more deterministic gene altering.

ML in Consumer Genomics

As the cost of genomic sequencing has plummeted, companies like ancestry.com and 23andme.com have popped up to provide "genome in a box" solutions to the general public. With millions of consumers submitting their genetic material, these companies employ ML to discover common patterns that can provide clues to ancestral origins, as well as predict weight gain, hair loss, and disease risk.[11]

---

[11] https://www.23andme.com/en-ca/dna-health-ancestry/

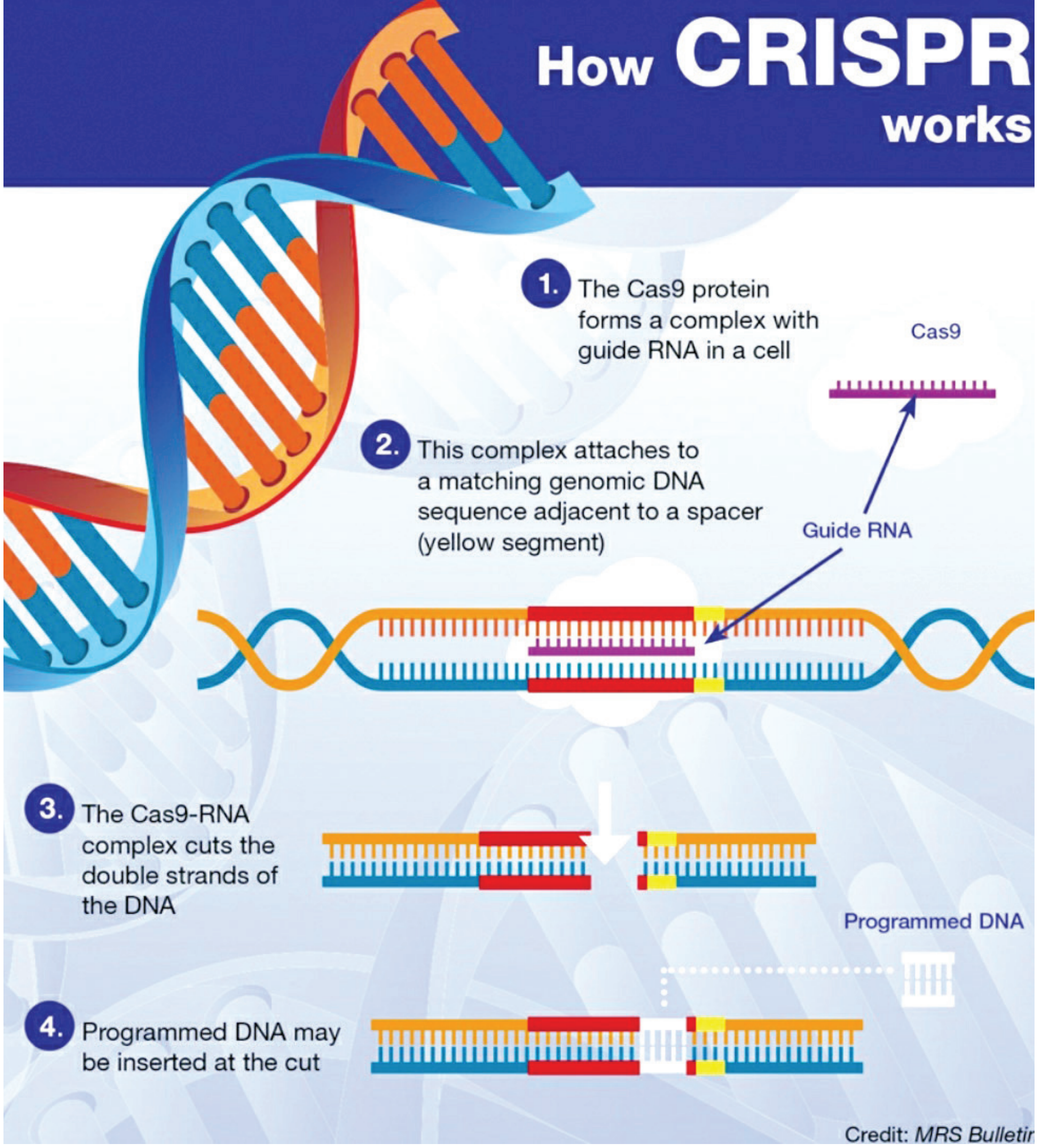


FIGURE 6: HOW CRISPR WORKS

ML IN GENE EXPRESSION

ML has been used to analyze gene expression data for decades. Unlike static sequence data, gene expression is a dynamic snapshot of genes in action.

Given that the human genome is composed of up to 20,000 genes, it is is an extremely difficult problem to determine which genes get expressed based on a limited amount of collected data.

When a gene is expressed in a cell it gets translated into messenger RNA (mRNA). Using micro-array technology researchers are able to measure the quantity of mRNA associated with a given gene sequence at any point in time. As a result, researchers can test, for example, cancerous vs non-cancerous tissue in order to measure how active/inactive specific genes may be in each sample.

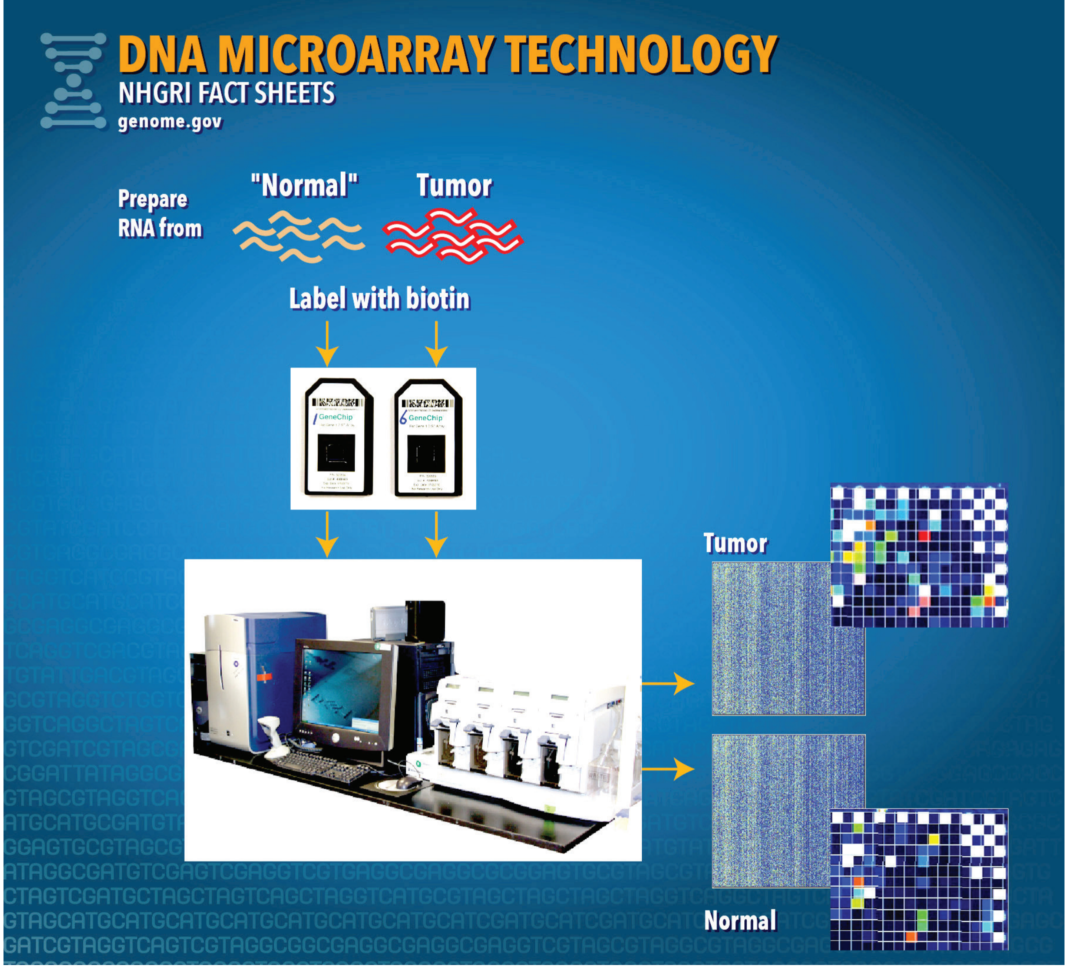


FIGURE 7: MICRO-ARRAY METHODOLOGY

While modern medical research is able to associate some genomic markers with disease risk, except in a few cases, these associations are not predictive. The problem lies in the fact that most diseases have multiple underlying genetic markers, some of which may be interacting in complex ways with the environment. Add to this the fact that the most significant signals in the data are often completely suppressed by technical noise, and the problem becomes even harder.

However, the future is promising. As ML has become more mainstream, introducing new algorithms and model building techniques to the community, the ability to identify patterns within complex gene interactions has improved. As a result, we may soon be able to treat genetic disorder causes rather than just their symptoms.

ML in Agricultural Genetic Research

CRISPR's gene-editing technology is being used to research the creation of high-yielding, drought-tolerant crops. But lots of genomic regions are similar. So CRISPR can edit the wrong gene sequence & result in unintended consequences, or "off-target effects." A collaboration between Microsoft and universities has created software tooling called "Elevation" and "Azimuth" (available on Microsoft Azure). It uses ML to predict on-target and off-target effects, improving CRISPR success rates.

"The holy grail in health care is not fancier technology and tools, it is physician and patient behaviour change. Machine learning will truly come of age when it can improve the decision-making of clinicians and patients or improve their efficiency in carrying out the actions that follow from those decisions."

Jean Drouin, M.D.

CHAPTER 5

DIAGNOSTICS

Medical diagnostics are a category of medical tests designed to detect infections, medical conditions and diseases. Diagnostic errors have been attributed to as much as 10% of all patient deaths, and may also account for between 6 and 17% of all hospital complications. [12]

ML has been seen as one potential solution to diagnostic challenges, particularly when applied to image recognition in oncology (e.g., cancer tests) and pathology (e.g., bodily fluid tests). In addition, ML has also been shown to provide diagnostic insights when examining Electronic Health Records (EHR).

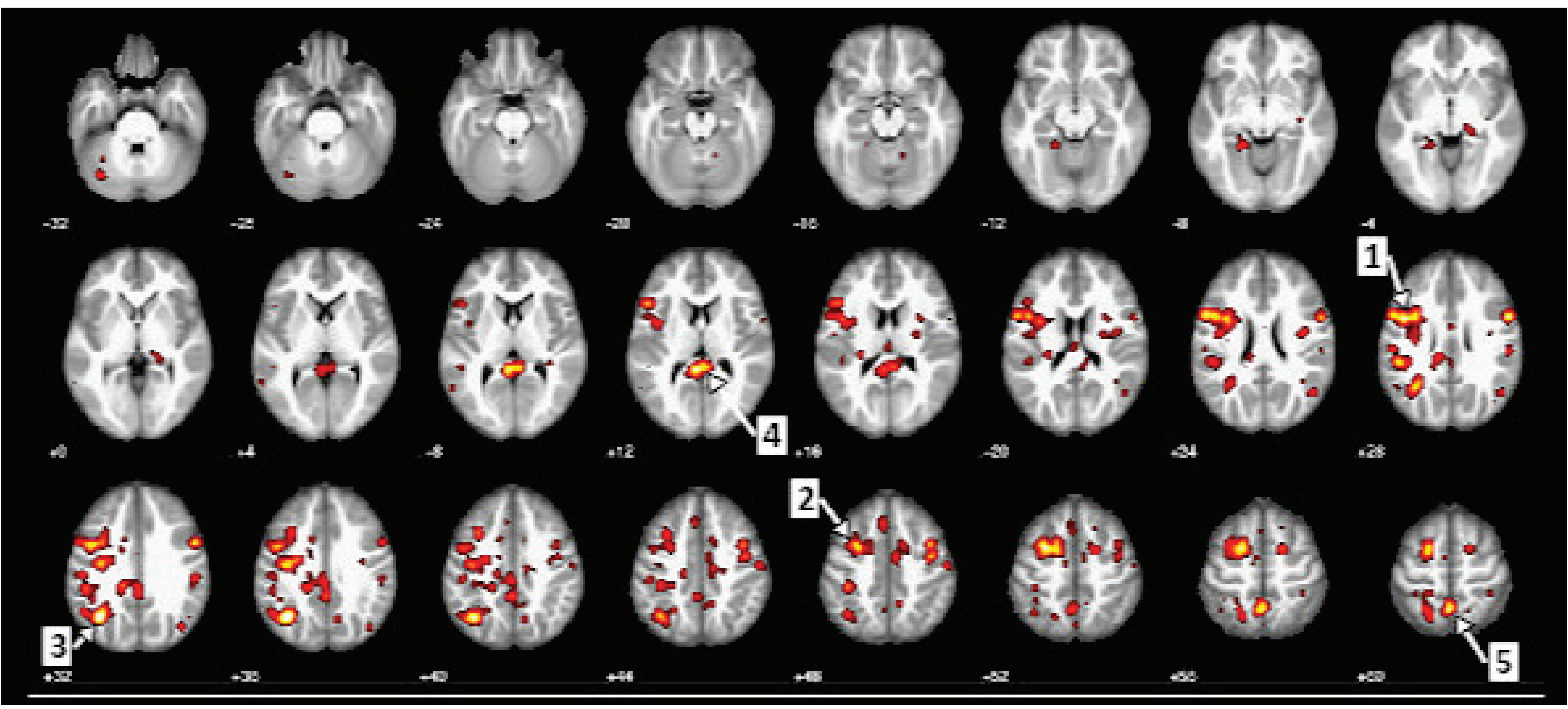


FIGURE 8: IBM-MIT PARTNERSHIP USES ML TO HELP DOCTORS PERFORM DIAGNOSES

[12] http://www.nationalacademies.org/hmd/~/media/Files/Report%20Files/2015/Improving-Diagnosis/DiagnosticError_ReportBrief.pdf

MEDICAL IMAGING

Image analysis and image-based diagnosis are two of the oldest areas of healthcare research to be targeted by ML. Getting sufficiently large datasets that are correctly classified is the key requirement for ML-based imaging. The good news is there are a number of publicly available imaging datasets,[13] and that number is growing every day. Unfortunately, differences between imaging systems remains an issue: a model trained on an imaging system from one manufacturer may not perform as well on images from another manufacturer.

At the end of the day, ML models still need to compete with senior clinicians, who (at least for now) routinely provide better results. At the other end of the spectrum, however, ML solutions can definitely challenge the performance of a junior clinician. For example, pathology studies have shown ML success rates of up to 92% compared to senior clinician success rates of 96%.

However, best results are achieved when combining ML with clinician results. For example, when ML diagnoses are vetted by pathologists, an accuracy rate of 99.5% is achieved.[14]

In oncology studies of skin cancer, where the first step is a visual examination of the patient's skin, initial studies show that ML can deliver results on par with professional dermatologists.[15]

More promising is the use of ML to provide diagnoses based on multiple modal inputs, such as Computerized Tomography (CT), Magnetic Resonance Imaging (MRI) and Diffusion Tensor Imaging (DTI) scans. While each scan provides a different way to evaluate an anomaly, human diagnoses tend to be less accurate than ML solutions since the human eye and visual cortex have a difficult time integrating radically different views into a single unified whole. By contrast, ML solutions have been shown to better be able to process each unique piece of information into a single diagnostic outcome.[16]

[13] http://www.aylward.org/notes/open-access-medical-image-repositories
[14] https://hms.harvard.edu/news/better-together
[15] https://news.stanford.edu/2017/01/25/artificial-intelligence-used-identify-skin-cancer/
[16] https://www.hindawi.com/journals/bmri/2017/1278329/

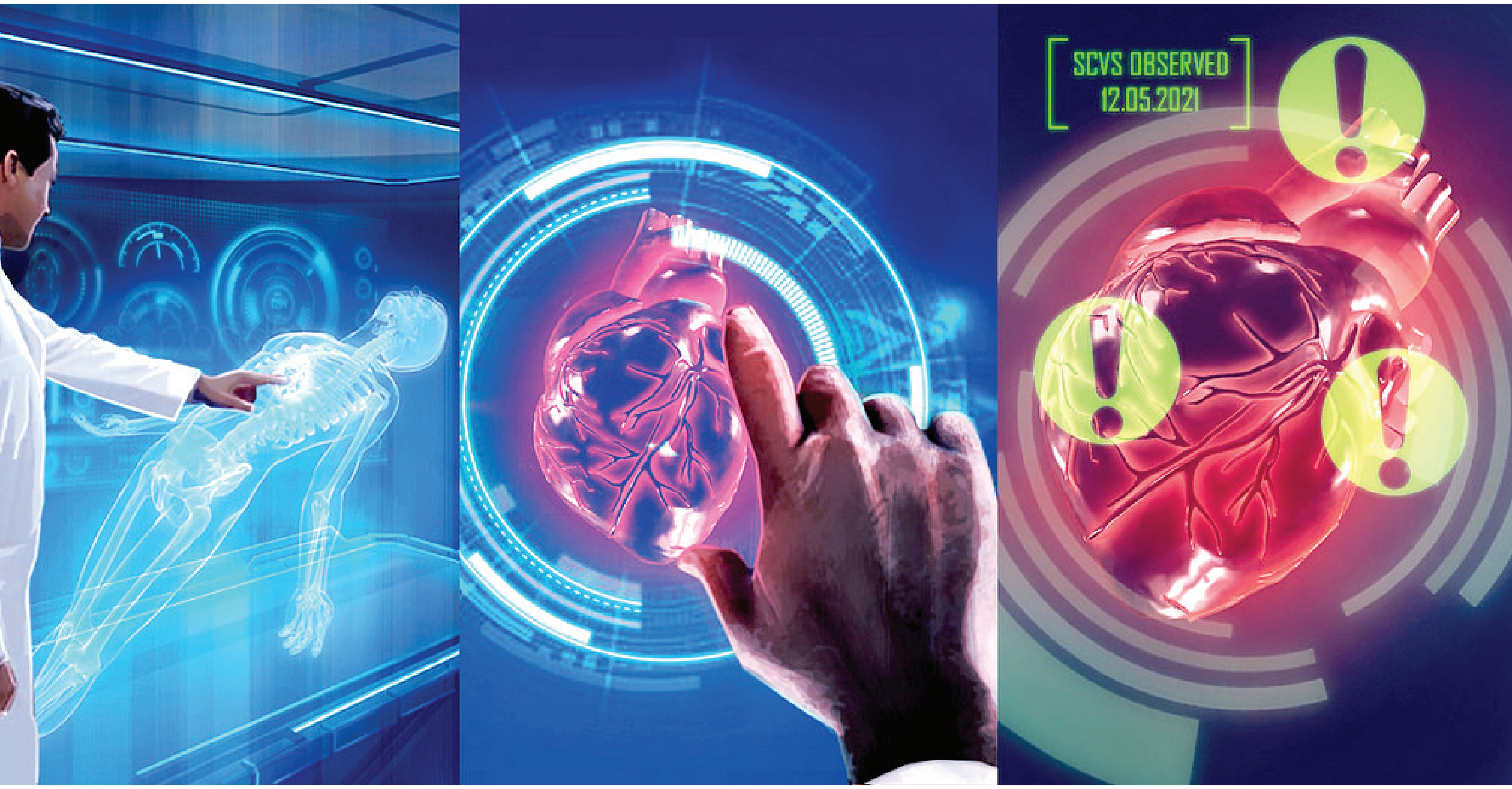


FIGURE 9: FUTURISTIC MULTI-MODAL IMAGING

Outcome prediction is an area in which clinicians often fall short, leading to ineffective management choices that raise costs. Imaging provides physicians with prognostically-valuable data but predicting disease progression requires the examination of multiple features interacting at differing scales, a feat difficult for humans to correctly determine with good reliability. As such, the role of ML in medical imaging will continue to grow, both as an aid to human diagnosticians, as well as a predictor of outcomes.

ML in Rare Diseases

Many diseases can be identified by a simple visual examination. In fact, rare genetic disorders that are expressed as facial dysmorphic features can often be spotted in photographs. FDNA is one company that has capitalized on this fact, creating an ML-based facial recognition application called Face2Gene. Uploading a photo to Face2Gene kicks off the ML process that evaluates the photo against a library of phenotypes and provides best match syndromes, improving diagnosis. [17]

PATIENT RECORD DIAGNOSES

Heart disease is the number one cause of death worldwide. Today, heart disease predictions are based on applying a fairly simple algorithm to eight key risk factors including age, cholesterol level and blood pressure. However, these simple algorithms do not take into account influencing factors like medications, multiple disease conditions, and other non-traditional biomarkers.

When more complex ML algorithms were applied to a dataset of over 300,000 patient records to predict an individual's chance for a heart attack or stroke within a 10-year time horizon, results showed that the neural network model was significantly better at predicting cardiovascular disease[18]. Accurately identifying patients at high risk early allows for intervention by doctors to prevent serious events like cardiac arrest and stroke.

More surprisingly, ML has also proven critical to preventing suicides, which claims as many as 45,000 lives annually in the US alone. Traditional suicide prediction research focuses on factors like depression, stress and substance abuse but hasn't resulted in an accuracy rate much better than random guessing.

[17] https://www.face2gene.com/
[18] https://www.nottingham.ac.uk/news/pressreleases/2017/april/artificial-intelligence-can-accurately-predict-future-heart-disease-and-strokes-study-finds.aspx

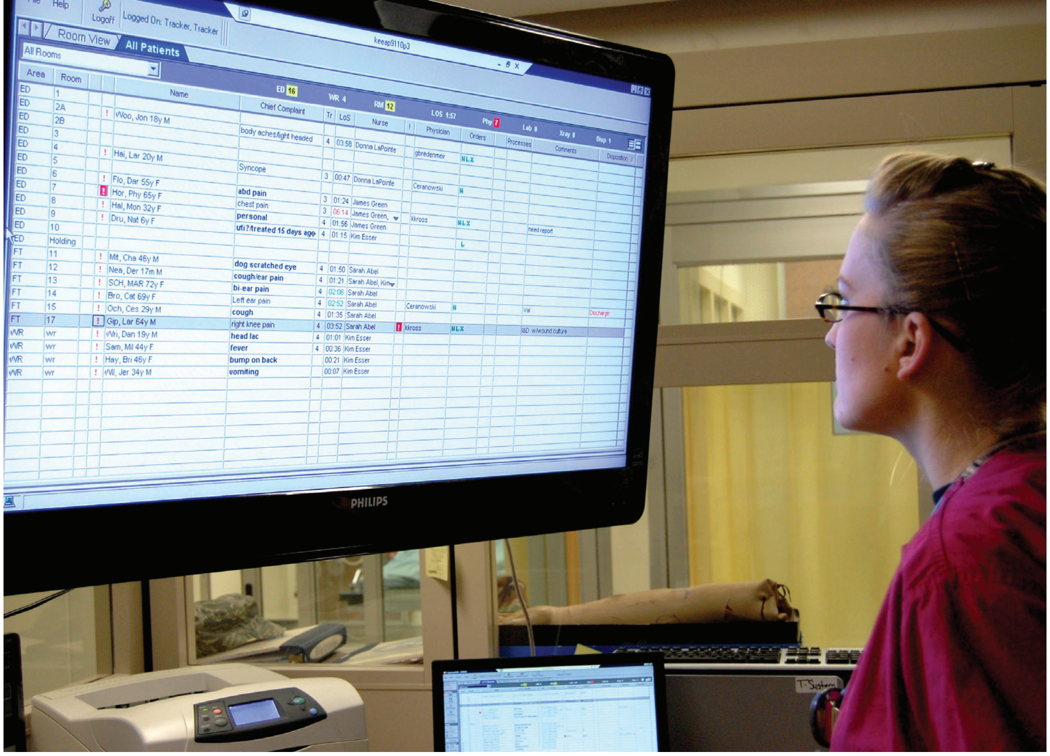


FIGURE 10: ELECTRONIC HEALTH RECORDS

A recent research project out of Florida State University[19] applied ML to the anonymous health records of 2 million patients in Tennessee. Having trained their model using the records of 3,200 patients who attempted suicide, the research team was able to apply algorithms to hundreds of data points in a patient's medical record, and create a suicide risk score that can predict with 80-90% accuracy whether someone will attempt suicide within two years. When the time horizon was restricted to just one week, the ML model was able to predict suicide attempts with a 92% accuracy for general hospital patients.

ML, whether applied to EHR or medical images, promises to help cut costs by replacing or supplementing clinical diagnostic work, while helping to deliver quicker, more accurate diagnoses. And since early diagnosis is key to reducing overall healthcare costs and improving outcomes for most diseases, ML usage is poised to grow at a rapid pace going forward.

ML in Medical Research

Founded in 2013, BenevolentAI employs ML to "read, analyze and understand" the vast amounts of scientific papers, patents, clinical trials, data, and images available in both public and private domains in order to help researchers speed up medical discoveries. Based on the data, it then creates hypotheses about likely cures for various diseases and has already made major breakthroughs, including one related to ALS.[20]

[19] https://medicalxpress.com/news/2017-03-machine-suicide.html
[20] http://benevolent.ai/blog/benevolentai/what-if-ai-could-take-your-research-to-the-next-level/

"AI, when used for medical imaging diagnosis and streamlining hospital workflows can enhance healthcare outcomes by 30 to 40% while reducing the cost of treatment by close to 50%."

Harpreet Singh Buttar, Frost & Sullivan analyst

Fast track your ML projects with
ActiveState's Python packages and tools.

Contact us to find out why 97% of
Fortune 1000 use our software:

solutions@activestate.com

ActiveState®

website: www.activestate.com | Toll-free in NA: 1.866.631.4581 | email: solutions@activestate.com