



# First Look

The Next Wave of AI Breakthroughs  
in Health Care

WORLD MEDICAL  
**INNOVATION**  
F O R U M





## 2017 First Look | The Next Wave of Cardiovascular Breakthroughs

Characterizing an Early HFpEF Phenotype: Cardiometabolic Disease and Pulmonary Hypertension

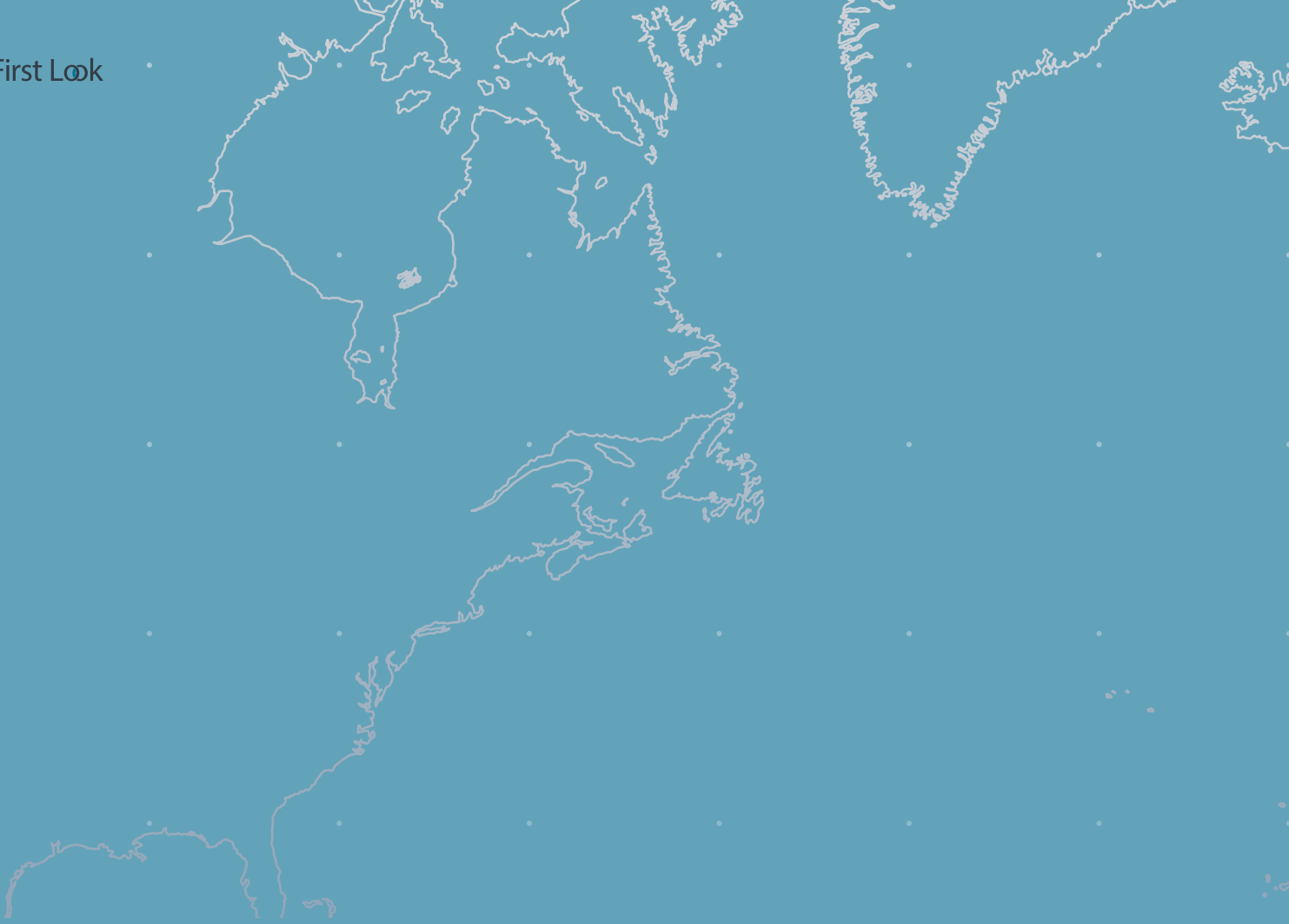
Jennifer Ho, MD

Assistant Physician in Medicine, MGH; Member of the Faculty of Medicine, Harvard Medical School

---

Early career Harvard Medical School investigators kick-off the 2018 World Medical Innovation Forum with rapid fire presentations of their high potential new technologies. Nineteen rising stars from Brigham Health and Massachusetts General Hospital will give ten-minute presentations highlighting their discoveries and insights that will disrupt the field of artificial intelligence. This session is designed for investors, leaders, donors, entrepreneurs, investigators and others who share a passion for identifying emerging high-impact technologies.

The top presenter each from BWH and MGH will be awarded the Peter K. Ranney Innovation Award. The prize carries a \$10,000 award.



**peter k. ranney  
innovation award**

The Peter K. Ranney Innovation Award will be given April 25<sup>th</sup> to honor a single BWH and single MGH First Look presenter who embodies the innovative, entrepreneurial and visionary spirit that the World Medical Innovation Forum was established to recognize. The two \$10,000 awards will be granted based on overall presentation quality, innovativeness, commercial potential, caliber of disruption, and market need. The Award will be judged throughout the morning session with winners announced at the annual Innovator's Dinner on Wednesday, April 25, 2018.



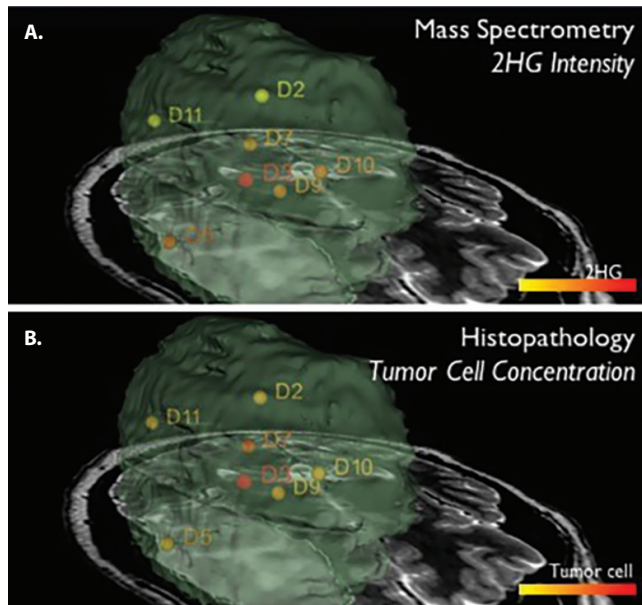




2018  
artificial intelligence  
investigators

- 6 Nathalie Agar, PhD
- 7 Omar Arnaout, MD
- 8 Jason Baron, MD
- 9 David Craft, PhD
- 10 Raúl San José Estépar, PhD
- 11 Andrey Fedorov, PhD
- 12 Georg Gerber, MD, PhD
- 13 Daniel Hashimoto, MD
- 14 Kasper Lage, PhD
- 15 William Lane, MD, PhD
- 16 Quanzheng Li, PhD
- 17 Maulik Majmudar, MD
- 18 Thomas McCoy, MD
- 19 Ziad Obermeyer, MD
- 20 Michael Rosenthal, MD, PhD
- 21 Hadi Shafiee, PhD
- 22 Erica Shenoy, MD, PhD
- 23 Brandon Westover, MD, PhD
- 24 Sabine Wilhelm, PhD

*Note: Speakers and content are subject to change.*



**Figure A.** 3D mapping of 2-HG detected by MS over MRI volume reconstruction.

**Figure B.** Histopathology scoring of tumor cell concentrations determined from H&E-stained tissue sections.

Santagata, S. et al (2014) PNAS. 111(30):1121-6. <http://www.slicer.org/>

## Mass Spectrometry Imaging for Surgical Pathology and Oncology



### Nathalie Agar, PhD

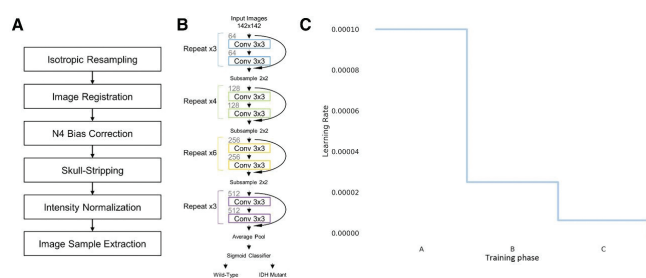
Research Scientist, Neurosurgery, BWH; Associate Professor, HMS

[Nathalie\\_Agar@dfci.harvard.edu](mailto:Nathalie_Agar@dfci.harvard.edu)

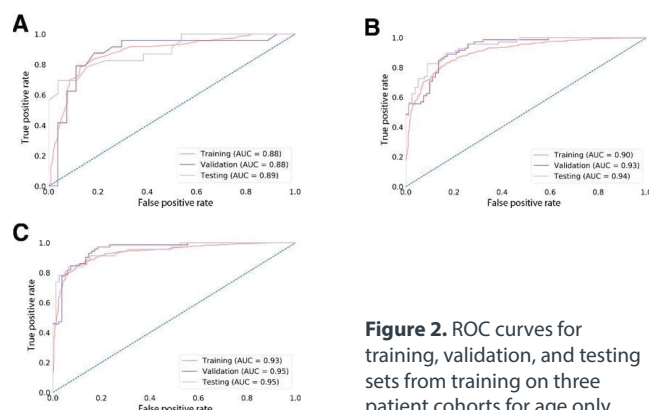
Mass spectrometry provides multiple options for the direct biomolecular characterization of tissue to support clinical decision-making and provides significant insight in the development of drugs targeting tumors of the central nervous system. Using an array of mass spectrometry (MS) applications, we rapidly analyze tumor markers ranging from small metabolites to proteins from surgical tissue for rapid diagnosis and surgical guidance. Using similar clinical protocols, we visualize drug and metabolites penetration in brain tumor tissue and correlate with tumor heterogeneity and response to support drug development. The methodologies developed and validated include the direct and quantitative analysis of specific biomarkers, and the application of machine learning algorithms to the complex data to identify diagnostic patterns. More specifically, we quantitatively monitor the distribution of the oncometabolite 2-hydroxyglutarate as a marker of the presence of cancer cells in glioma tumors harboring a mutation in the isocitrate dehydrogenase (IDH) gene and provide the information in real-time to the neurosurgeon to further inform surgical decision making. Using the

broader mass spectral information containing hundreds to thousands of peaks, we apply classification workflows to distinguish brain tumor types, subtypes, grades, and assess tumor cell concentration, and continue to investigate different AI tools to facilitate analysis and extract more information. In an effort to increase our specificity in delineating pituitary microadenomas, we combine the analysis of specific pituitary hormones with machine learning. Furthermore, we apply similar approaches to assess tumor margins for breast conserving surgery in the treatment of breast cancer.

While we have made this technology available and continue to further validate it at BWH, true dissemination of the work will occur via commercialization of both clinically compatible mass spectrometry systems, and well validated human reference databases and algorithms allowing the rapid acquisition, analysis, and visualization of diagnostic classification.



**Figure 1.** A. Image preprocessing steps. B. Modified 34-layer residual neural network architecture. C. The learning rate schedule.



**Figure 2.** ROC curves for training, validation, and testing sets from training on three patient cohorts for age only.

## Using Residual Convolutional Neural Networks to Improve Treatment of Brain Tumors



### Omar Arnaout, MD

Co-Director, Computation Neuroscience Outcomes Center, Attending Neurosurgeon, Department of Neurosurgery, BWH; Member of the Faculty, HMS

[oarnaout@bwh.harvard.edu](mailto:oarnaout@bwh.harvard.edu)

Gliomas are the most common primary brain tumor in adults and the cure remains elusive. Certain pathological biomarkers including the mutation status of isocitrate dehydrogenase (IDH) have significant implications on the benefits of aggressive surgical resection, response to treatment and overall prognosis. Tissue sampling via biopsy or surgical resection has been required to elucidate biomarker status.

Our research is designed to predict IDH status of gliomas using preoperative structural MR imaging by utilizing a residual convolutional neural network.

Preoperative imaging for patients with grade II-IV gliomas was obtained from multiple institutions including Brigham and Women's Hospital, the Hospital of the University of Pennsylvania and The Cancer Imaging Archive. The dataset was divided into training, testing and validation sets. A residual convolutional neural network was trained from each MR sequence to build a predictive model. Data augmentation in the form of image rotation, translations, flips, shearing and zooming was utilized. The age at the time of diagnosis was also incorporated into the model.

The trained neural network achieved IDH prediction accuracy of 87.3% (AUC 0.93), 87.6% (AUC 0.95), and 89.1% (AUC 0.95) with training, validation and training sets respectively.

A multi-institutional dataset was used to generate and validate a deep learning model capable of predicting with high accuracy IDH mutation status in patients with low and high grade gliomas. As a result, we now have a tool that can be deployed in the clinical setting allowing us to better counsel patients, and tailor the surgical approach and goal to the individual. Furthermore, this project serves as a proof-of-concept for a suite of tools that will use deep learning to predict various biomarkers in patients with brain tumors, augmenting traditional radiological interpretation for clinical decision making.

# Machine Learning-Based Clinical Decision Support for Laboratory Test Interpretation



## Jason Baron, MD

Medical Director, Core Laboratory, MGH;  
Assistant Professor, HMS

[jmbaron@partners.org](mailto:jmbaron@partners.org)

Many clinical decisions rely on laboratory test results. An individual patient may have laboratory results from tens or hundreds of tests across a range of time points, requiring a clinician to manually assimilate and interpret many data points to inform diagnosis, prognosis or treatment. This type of manual interpretation of test results will not only consume a clinician's scarce time, but will often fail to maximally extract clinical information, particularly given the many complex interrelationships between tests, some of which likely have yet to be discovered. Our group seeks to advance the ways in which clinical laboratories report test results by applying machine learning to real world clinical data to discover new knowledge and build models to automatically generate patient-specific interpretations of clinical laboratory data.

For example, we had developed, in collaboration with computer scientist colleagues, a machine learning-based algorithm that could predict ferritin test results (a marker of iron status used in the diagnosis of iron deficiency) using only patient demographics and results from other tests performed on the same patient. We demonstrated not only that this algorithm could predict ferritin results suggestive of iron deficiency with a high degree of accuracy (AUC=0.97), but that predicted ferritin test results were in many cases more informative than measured ones. This algorithm has several potential applications. Laboratories could automatically calculate a "predicted ferritin" alongside every measured ferritin; measured ferritin values differing substantially from predictions could be reported with an interpretive comment to reduce the risk of iron deficiency misdiagnosis. Likewise, we could potentially adapt the algorithm to identify patients in whom a ferritin test was not ordered but who might have iron deficiency or could benefit from ferritin testing. We have applied similar approaches to other laboratory tests and anticipate that the ferritin approach is generalizable to a wide-range of applications. Our group is also developing strategies to implement into clinical practice machine-learning based clinical decision support algorithms, such as the ferritin prediction algorithm.

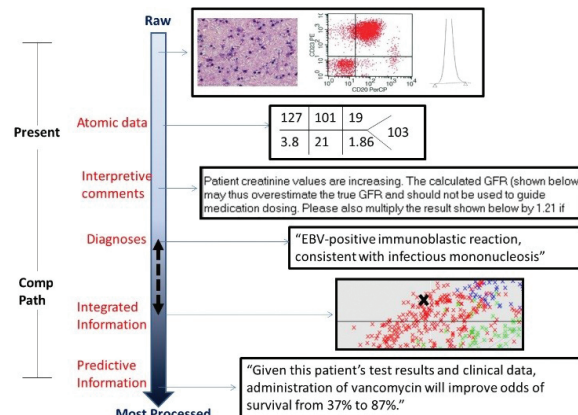


Figure 1. Schematic of big picture goals.



Figure 2. Raw data used for the ferritin prediction.

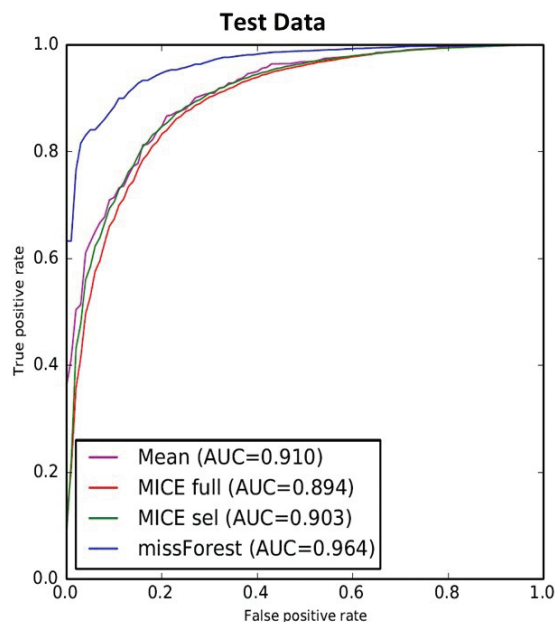


Figure 3. Overview of ferritin prediction methods and results.





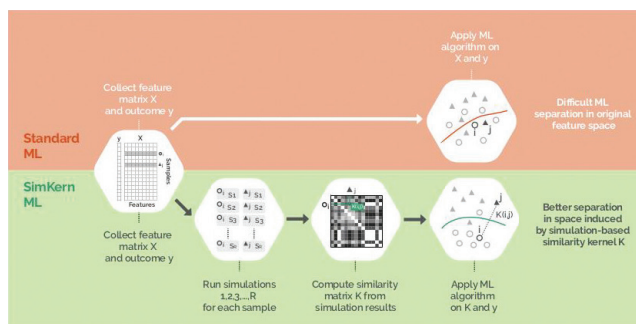
# Leveraging Machine Learning for Personalized Cancer Treatments

**David Craft, PhD**

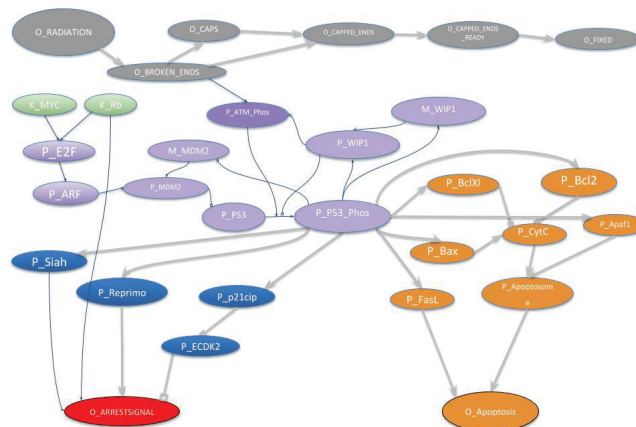
Staff Scientist, MGH; Assistant Professor, Radiation Oncology, HMS

[dcraft@broadinstitute.org](mailto:dcraft@broadinstitute.org)

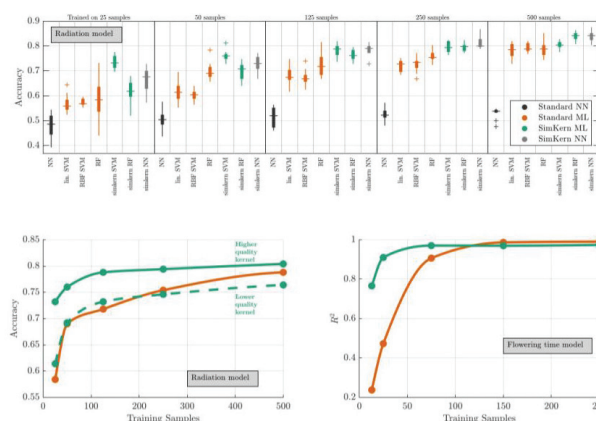
It is not obvious how decades of hard won biological knowledge can be incorporated into building machine learning algorithms for predicting how a drug will affect a patient. We study this problem from theoretical, computational, and practical angles. Our computational studies provide strong evidence that incorporating system knowledge, rather than using pure data driven machine learning approaches, offers substantial improvements to predictive accuracy. We describe open source machine learning datasets we have created for studying methods to incorporate prior knowledge and to assess the value in doing so. Boolean networks, graphs of interconnected logical (AND, OR) switches, represent complex dynamical systems and are used by systems biologists to model cellular signaling. We produce large random Boolean networks and compare machine learning methods which use knowledge of the underlying network connectivity with those that do not. We also use simulation models of biological processes (some taken from literature [e.g. a flowering time prediction problem], others created by our group [e.g. cellular response to DNA damage from radiation]) to generate datasets. We demonstrate in all cases that incorporating prior knowledge significantly enhances predictive capacity. Turning to real datasets, we present ideas and results for machine learning a cell line radiation sensitivity experiment. Prior knowledge here takes the form of expert gene selection, automated PubMed searches, Watson for Drug Discovery, and/or incorporating hierarchical biological information available at geneontology.org. We argue that such knowledge incorporation is critical given the “large p small n” regime we are in: p=number of (genetic) parameters, 100s of thousands, and n=number of samples we have, usually in the 100s. The personalized cancer medicine problem is in its infancy due to the complexity of human cancers. This talk will reflect on what makes this problem so difficult and will give evidence that multidisciplinary efforts to include existing biological knowledge are of vital importance to developing a high quality clinical prediction tool.



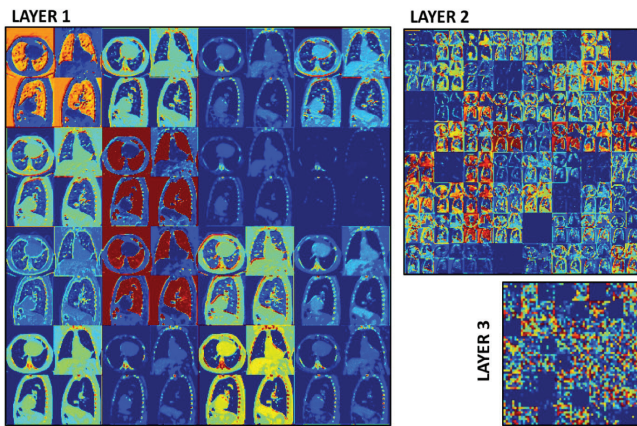
Example technique to build in prior knowledge via detailed simulations. ML=machine learning.



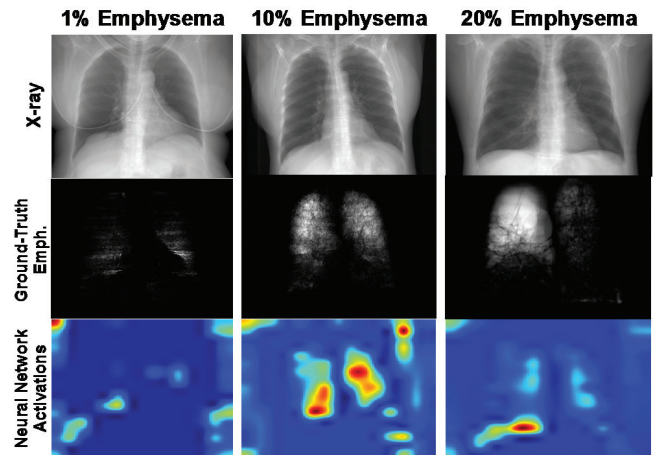
Cellular response to radiation model, used to generate data for machine learning algorithm testing.



Results demonstrating superiority of prior knowledge (here called SimKern, for simulation based kernel learning) machine learning. NN=nearest neighbor, SVM=support vector machine, RF=random forest, RBF = radial basis function. Accuracy is classification accuracy,  $R^2$  is coefficient of determination.



**Figure 1.** Response of the mortality prediction convolutional neural network for a test subject.



**Figure 2.** Quantification of Emphysema from X-Ray images and corresponding network activations.

## Identification of Healthcare Risks in COPD Patients Using CT and X-ray Images



### Raúl San José Estépar, PhD

Co-Director, Applied Chest Imaging Laboratory, BWH; Associate Professor, Radiology, HMS

[rsanjose@bwh.harvard.edu](mailto:rsanjose@bwh.harvard.edu)

Chronic obstructive pulmonary disease (COPD) is one of the most common respiratory conditions that affects more than 22 million Americans; it has climbed to be the third leading cause of death and accounts for \$32 billion in associated healthcare cost. COPD remains underdiagnosed, and its associated comorbidities include cardiovascular disease, musculoskeletal disorders, and lung cancer. Proper population management of COPD patients rests on the ability to define who is at a higher risk based on their disease prognosis and underlying comorbidities.

Chest Computed Tomography (CT) and X-ray (CXR) imaging provide a unique snapshot of patients' health status that includes three major systems: heart, lung and skeletal muscle. I will present the latest advancements that my group has developed in image-based artificial intelligence approaches to providing risk assessment of patients with COPD. Our risk models are based on deep learning approaches to quantify emphysema, air trapping, heart size, body composition and bone mineral density from both chest CT and CXR images. We have also developed prognostication models based on canonical landmark views of a CT scan that can stage and prognosticate acute respiratory event and death in COPD.

With annual utilization rates of CT and CXR of 400 and 900 exams respectively per 1,000 beneficiaries just in the Medicare population alone, the deployment of these tools to mine standardized Picture Archiving and Communications Systems (PACS) can enable healthcare analytic companies with population management solutions based solely on imaging information.



# Data Standardization for Advancing the Role of AI in Radiology

## Andrey Fedorov, PhD

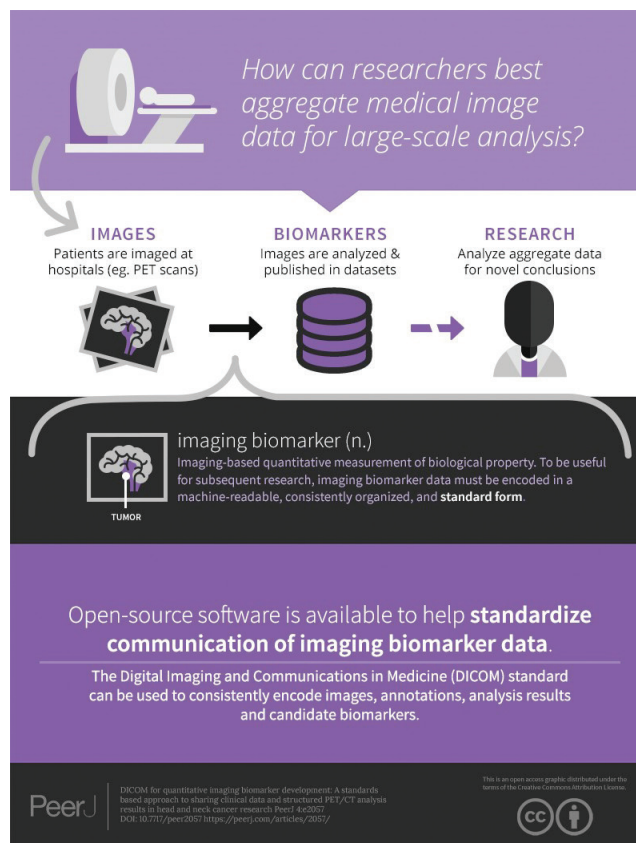
Assistant Professor, Radiology, BWH; Assistant Professor, HMS  
[fedorov@bwh.harvard.edu](mailto:fedorov@bwh.harvard.edu)

Success of an AI system depends on the amount and quality of data used to train it. The database that was key to the latest AI revolution (ImageNet) contains millions of real-life images labeled into thousands of categories. No databases of comparable extent and quality exist for radiology data. By many, this is considered to be the biggest challenge for AI in radiology.

Training of AI models requires medical images accompanied by expert annotations (e.g., spatial location of the finding, its clinical characteristics), linked with the non-imaging part of the patient record. While large volumes of clinical images are routinely collected and interpreted both in clinical and research studies, the result is often not reusable. Open standards have been implemented broadly to alleviate some of the technical challenges of interoperability for medical images. No solutions were adopted however for standardization of the image-derived data. Standardization of the routinely collected manual and automatic annotation results opens dramatic opportunities for the development of AI tools and improved understanding of the disease.

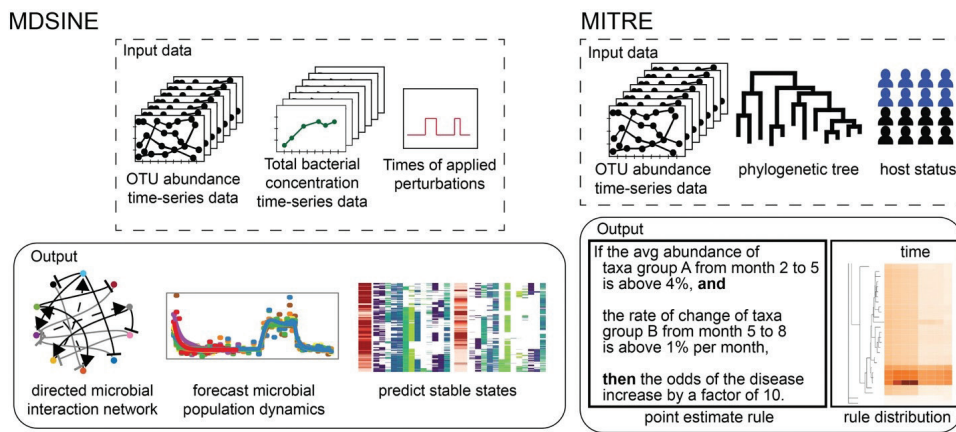
Our goal is to incrementally advance standardization in computational analysis of medical images. We are extending the existing, broadly adopted DICOM standard, to support the needs of medical imaging research applications, and subsequent implementation into existing clinical systems. We developed open source tools that support standardization of common outputs of image analysis, a functionality that did not exist in either academic or commercial products. We established collaborations with a number of academic and industry groups to support and evaluate adoption. With the increasing adoption of the standard, and lack of viable alternative solutions, the project is beyond the pilot stage. Applications in large scale analysis initiatives and adoption in mainstream production tools will be critical to advance and fully leverage the power of AI in radiology.

Universal adoption of image standards by the equipment manufacturers enabled dramatic growth and progress. Standardization of image-derived data can enable the AI revolution in radiology.



**Figure 1.** Graphical abstract of a related manuscript covering some of the details of the topic, available at <https://peerj.com/articles/2057/>





**Figure 1.** MDSINE and MITRE tools for developing microbiome therapeutics and diagnostics.

## Machine Learning Tools for Microbiome Diagnostics and Therapeutics Development



### Georg Gerber, MD, PhD

Associate Pathologist, BWH; Assistant Professor, Pathology, HMS  
[ggerber.bwh.harvard.edu](http://ggerber.bwh.harvard.edu)

There is evidence that the microbiome plays an important role in a variety of human diseases including infections, arthritis, food allergy, cancer, inflammatory bowel disease, neurological diseases, and diabetes. This evidence has sparked intensive R&D on diagnostics and bacteriotherapies or “bug-as-drugs.” There have been >45 disclosed equity investments in microbiome therapeutics since 2010, totaling > \$840M, with a projected market size of \$3.2B by 2024.

Thus, there is a growing need for robust and predictive computational models to accelerate diagnostic and bacteriotherapy development. The problem is challenging, because it involves modeling dynamic and complex microbial ecologies with patient-specific aspects.

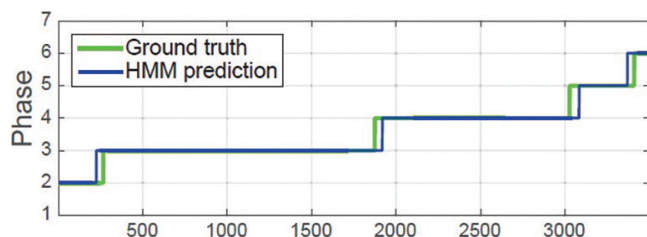
We have developed novel Bayesian machine learning tools for addressing two complementary problems in the field: (1) predicting status of the host (e.g., disease) from microbiome dynamics, to develop diagnostics, and (2) predicting population dynamics of bacteriotherapies and their interactions with the microbiome.

The Microbiome Interpretable Temporal Rule Engine (MITRE) is novel a Bayesian supervised learning method that automatically learns, from microbiome time-series data, probabilistic classifier models of host status. These classifier models consist of sets of human-readable rules about time-localized patterns of change in the abundances of groups of phylogenetically related organisms.

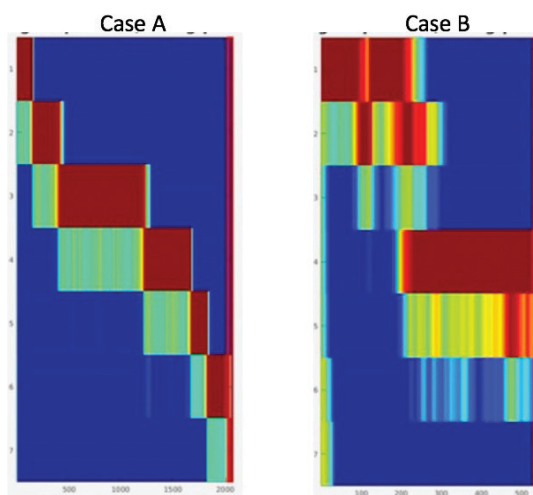
The Microbial Dynamical Systems INference Engine (MDSINE) infers dynamical systems models from noisy microbiome time-series data. Our tool introduced new Bayesian approaches that include modeling time-series of next generation sequencing data, adaptive smoothing of sparse/irregularly sampled data, and explicit learning of underlying microbe-microbe interaction networks. Our recent extension of MDSINE, to include automatically learned modular structure, was selected for an oral presentation at the premiere machine learning conference NIPS 2017.

Applications of MDSINE and MITRE have already yielded a provisional patent for a diagnostic to predict patients at risk for *Clostridium difficile* recurrence and a bacteriotherapy to prevent the infection, and a sponsored research agreement to analyze clinical trials of dietary interventions to alter the microbiota. Commercialization strategies include providing software-as-a-service to pharmaceutical and biotech companies, with co-development and royalty sharing models.





**Figure 1.** Over 90% accuracy of machine segmentation (blue) vs. human segmentation (green).



**Figure 2.** Fingerprints comparing A) routine sleeve gastrectomy vs. B) sleeve + lysis of adhesions.

## Surgical Fingerprints: Real-Time Analysis and Summarization of Intraoperative Events



### Daniel Hashimoto, MD

Edward D. Churchill Surgical Education and Simulation Research Fellow, MGH  
[dahashimoto@mg.harvard.edu](mailto:dahashimoto@mg.harvard.edu)

Intraoperative adverse events (IAEs), such as bowel/vascular injury, are estimated to occur in 2% of operations. IAEs can exact a toll on patient quality of life and costs, with average admission charges estimated to be 41% higher for patients who experience IAEs. Up to two-thirds of surgical errors occur intraoperatively, and 86% of these are secondary to cognitive factors such as failures in judgment, memory, or vigilance that lead to poor decisions. Currently, analysis of the intraoperative phase of care is limited to review of dictated operative reports that are notoriously incomplete. While video has been shown to be more accurate for identification of IAEs, manual review of video is costly and time-consuming.

We developed a machine learning approach to analyze laparoscopic video and generate identification and segmentation of operative steps. Using annotated video, we trained a support vector machine and a neural network to classify video frames into their respective operative steps. Hidden Markov models were used for segmentation of videos into operative steps with >90% accuracy (Fig. 1). We used coresets to leverage semantic summarization of video to increase the efficiency of segmentation. The cumulative log probability for each frame allowed for real-time estimation of deviation from an expected operative path and resulted in a “surgical fingerprint” that visually summarized potential areas of unexpected operative events (Fig. 2).

Real-time analysis of surgical video provides the foundation for intra- and post-operative clinical decision support to augment surgical decision-making. We plan to link population-based pre-procedure risk scores with patient-specific surgical fingerprints with the goal of preventing IAEs and identifying optimal, patient-specific post-procedural management, with the potential to recognize rare patient scenarios and seemingly common patients who may not fit into standard protocols. By unlocking intraoperative care as a quantitative data source to predict IAEs, complications, and readmissions, we expect to have an impact on daily clinical care by providing clinicians with an effective, data-driven tool for perioperative management.

Guy Rosman<sup>2</sup>, Daniela Rus<sup>2</sup>, Ozanan R. Meireles<sup>1</sup>

<sup>1</sup>Department of Surgery, Massachusetts General Hospital

<sup>2</sup>Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology



# Leveraging AI and Biological Networks to Functionally Interpret Vast Genetic Datasets

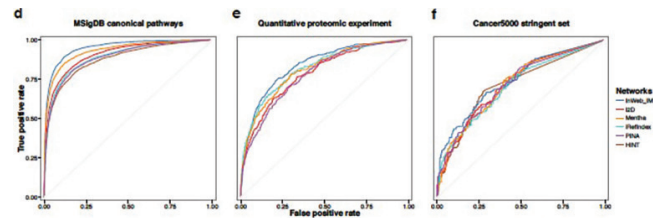
**Kasper Lage, PhD**

Director of Bioinformatics, MGH; Associate Professor, Surgery, HMS

[lage.kasper@mgh.harvard.edu](mailto:lage.kasper@mgh.harvard.edu)

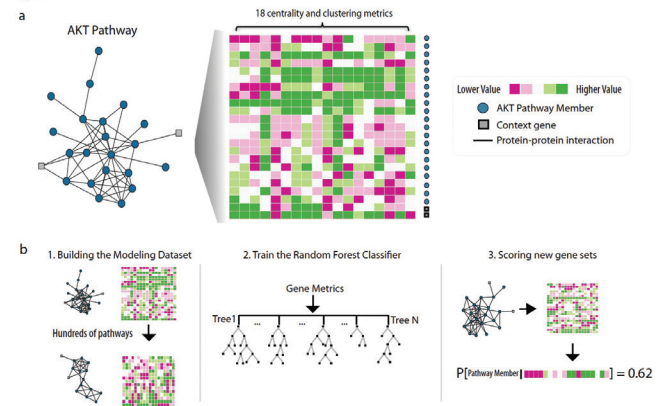
With the unprecedented ability to sequence genomes and to genotype millions of patients functional interpretation of genetic data is now one of the major bottlenecks in science and healthcare. This, in turn, significantly limits our efforts toward biological understanding and patient-specific therapeutic intervention in human diseases. Recently, high throughput technologies in transcriptomics and proteomics have provided insights into the wildly complex ‘social networks of genes’ by enabling us to measure the functional correlations, or physical interactions, between tens of thousands of genes in different tissues or cell types in a single experiment. Interpreting genetic datasets by finding unexpected functional connections in these networks has emerged as a powerful computational approach to identify unexpected biology with therapeutic and diagnostic value.

Towards these aims, we devised a computational framework to extract and quality control data from >40,000 scientific publications into a scored human protein–protein interaction network (InWeb\_InBioMap, Li et al Nature Methods 2017). We designed an artificial intelligence algorithm (Quack) that through a rigorous training procedure learns to identify unexpected pathway relationships in genetic data using any user-defined biological network. We leveraged these approaches to analyze genetic data from 4,742 cancer genomes to identify unexpected pathway relationships that predicted 62 new cancer driver candidates. In mouse experiments, we found that our candidate genes induce tumors at rates that are comparable to those of known oncogenes and confirmed our predictions by reanalyzing nine tumor-inducing candidates in 242 patients with oncogene negative lung adenocarcinomas. In this patient group, two genes (AKT2 and TFDP2) are significantly amplified [Horn et al, Nature Methods 2018]. To make our approaches widely accessible we developed a unified web platform, GeNets (<http://apps.broadinstitute.org/genets>) where users can upload proprietary biological networks and genetic datasets, train Quack models and execute, store, and share network analyses of genetic datasets [Li et al, accepted, Nature Methods]. Overall, we show how our technologies can interpret large genetic datasets and lead to therapeutic and diagnostic insights in patients with genetic diseases.

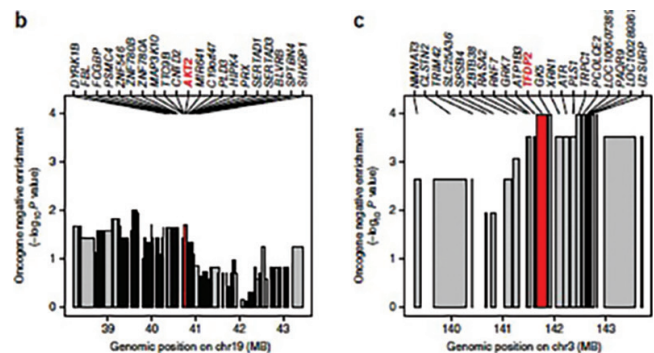


**Figure 1.** InWeb\_InBioMap has better functional biological relevance than other comparable resources.

Figure 2



**Figure 2.** Quack allows the identification of pathway patterns in genetic data.



**Figure 3.** Identifying new cancer driver genes in 4,700 cancer genomes.



# bloodTyper: Blood Group Typing using Next Generation Sequencing

**William Lane, MD, PhD**

Director of Clinical Lab Informatics, Assistant Director Tissue Typing Lab, BWH; Assistant Professor, HMS

[wlane@bwh.harvard.edu](mailto:wlane@bwh.harvard.edu)

Blood transfusion is one of the most common therapies administered in US hospitals, but sensitization to red blood cell (RBC) and platelet (PLT) antigens can lead to serious complications in prenatal medicine and transfusion. Although there are over 300 known RBC and 33 PLT antigens, current testing only includes matching the patient and donor for ABO and RhD antigens using methods that have not materially changed for over 60 years, despite not always distinguishing important changes.

Next generation sequencing (NGS), such as whole exome sequencing (WES) and whole genome sequencing (WGS), could improve antigen typing. However, without computerized algorithms capable of interpreting the NGS data, the translation to antigen phenotype is laborious, time-intensive, and requires deep subject matter expertise. To address this, we created a curated database of RBC and PLT antigen molecular changes (<http://bloodantigens.com>), followed by development of an automated NGS antigen typing software called bloodTyper (<http://bloodantigens.com/bloodtyper>), which we intend to commercialize.

bloodTyper was initially validated using 110 WGS samples, with perfect concordance to conventional serologic typing. We have also developed a targeted NGS ABO assay and updated bloodTyper to distinguish between ABO subtypes, with implications for stem cell and organ transplantation. bloodTyper will also be adapted to analyze an inexpensive SNP array together with a leading institution and company. We are also planning a collaborative project to improve bloodTyper using 100,000s of results from several large-scale genomics projects.

bloodTyper could routinely be used at low cost on anyone undergoing genomic sequencing, which in the future could be everyone.

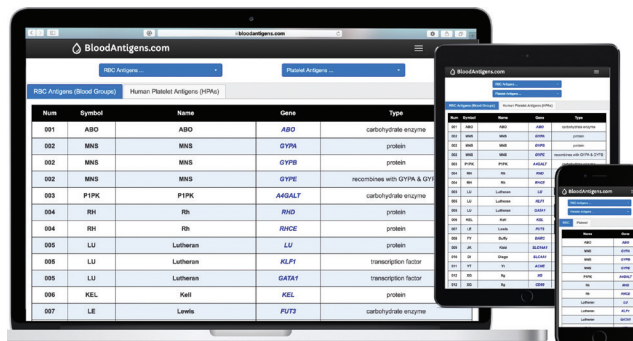


Figure 1.

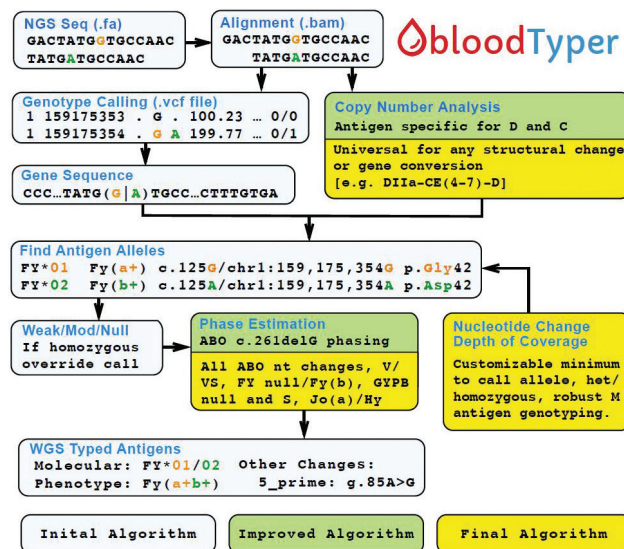


Figure 2.

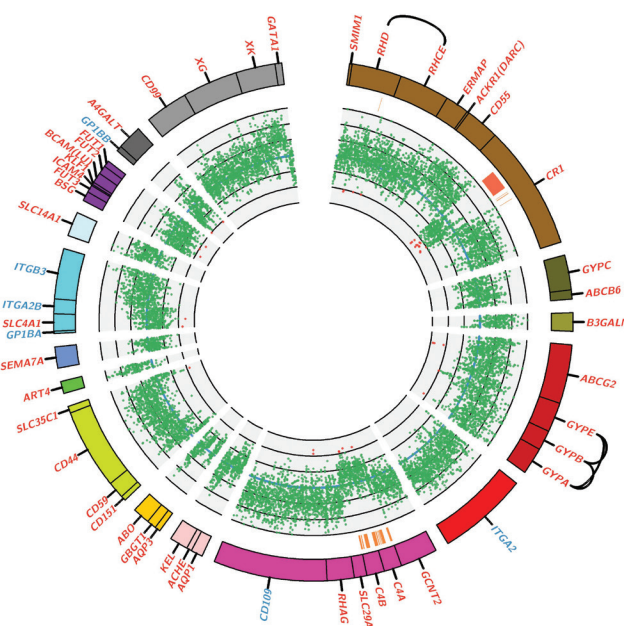


Figure 3.





## Deep Learning-Enabled System for Prescreening

### Quanzheng Li, PhD

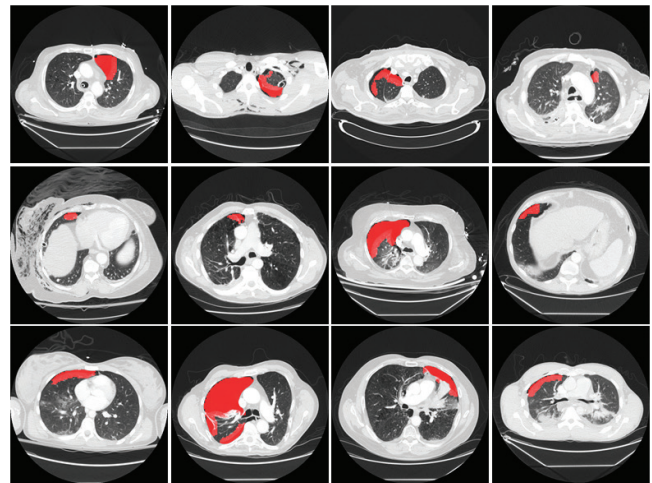
Director, Center for Advanced Medical Computing and Analysis, MGH; Associate Professor, Radiology, HMS

[li.quanzheng@mgh.harvard.edu](mailto:li.quanzheng@mgh.harvard.edu)

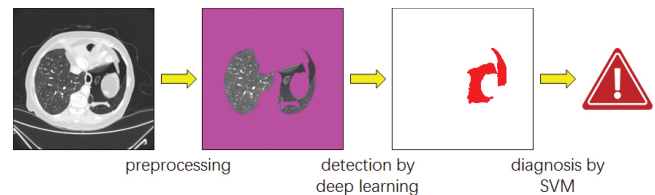
We developed a deep learning-enabled system for prescreening of critical conditions and demonstrate it using the detection and quantification of pneumothorax.

Our Institutional Review Board (IRB) approved study included transverse 3D CT images from 341 patients including 172 with pneumothorax (positive) and 169 without pneumothorax (negative). A 2-D patch-based convolutional neural network (CNN) was trained on 80 chest CT (50 positive and 30 negative cases) with a total of 4122 images. The training was repeated 10 times to create 10 individual networks. The CNN structure consisted of eight layers (five convolutional layers with kernel size of 2 followed by maximum pooling, and three fully connected layers). The network inputs were 36 x 36-pixel image patches, and the outputs were labeled patches. For each patient, the patch-wise results were aggregated into a lesion map depicting regions with pneumothorax. Image features such as size, relative location and shape descriptor of pneumothorax were extracted from the lesion map. A support vector machine (SVM), trained on these image features, was used to perform patient-wise prediction. Patch-wise accuracy of the CNN, and patient-wise accuracy of the SVM were determined.

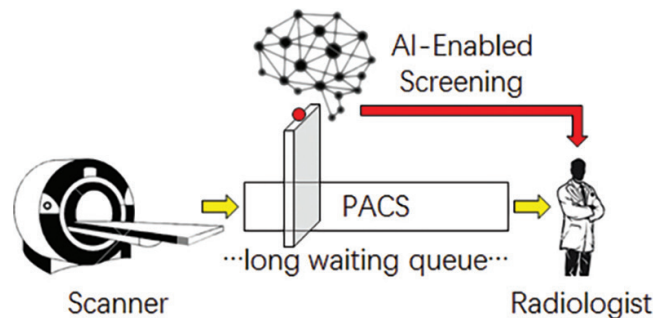
Our results show average patch-wise accuracy of the CNN model was 93%. Patient-wise accuracy for the test dataset was 94% with a sensitivity of 100% (all 122/122 pneumothorax detected) and a specificity of 89% (123 true negative/139 negative cases). The deep learning-enabled system has high sensitivity and accuracy for detection of pneumothorax on chest CT images and is ready to be extended to other critical conditions. We are further integrating the system into the clinical workflow at the point of completion of each chest CT scan, making it useful for patients with pre-existing chest diseases that place them at higher risk for pneumothorax, and for patients presenting in the emergency department for rapid triage.



**Figure 1.** Visualization of pneumothorax detection results (all positive) from nine patients, shown in axial view.



**Figure 2.** Illustration of analytics pipeline, from left to right: input CT image, lung region after preprocessing, pneumothorax region detected by deep learning, patient-wise diagnosis by SVM classifier.



**Figure 3.** Illustration of the need for a rapid screening system to re-prioritizing the work list for hospital management.



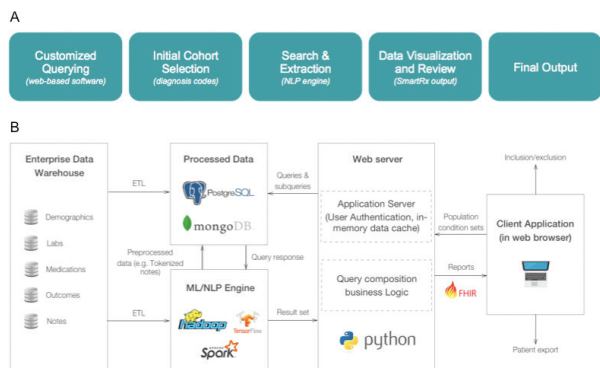


Figure 1: SmartRx Software Engine. A) Key components. B) System architecture

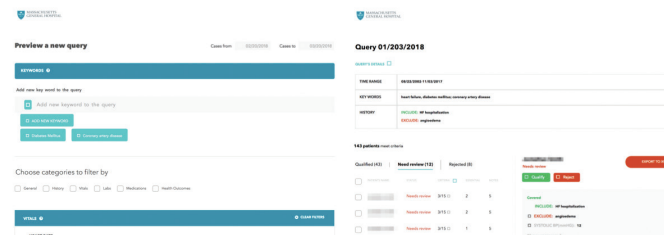


Figure 2: Representative screenshots of the SmartRx Software Platform.

## SmartRx: A Natural Language Processing-Based Software Platform that Enables Automated, Real-Time, Querying of the EMR

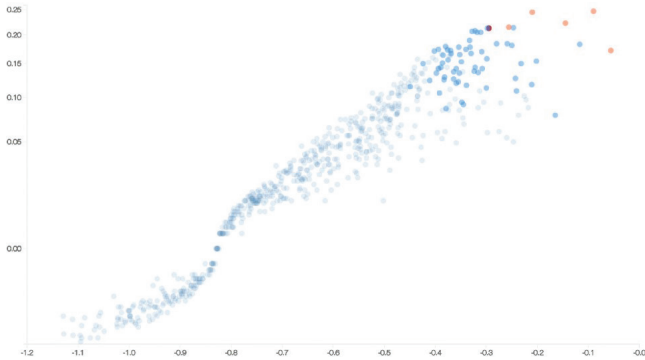


### Maulik Majmudar, MD

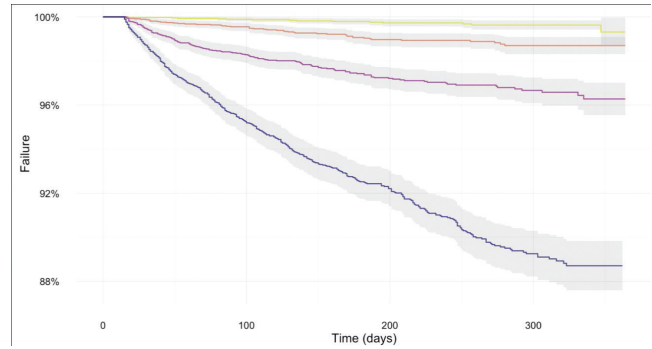
Associate Director, Healthcare Transformation Lab, MGH; Assistant Professor, Cardiology, HMS  
[mmajmudar@partners.org](mailto:mmajmudar@partners.org)

A learning health system is defined as a system in which, “science, informatics, incentives, and culture are aligned for continuous improvement and innovation, with best practices seamlessly embedded in the delivery process and new knowledge captured as an integral by-product of the delivery experience.” The first step in the journey towards a learning health system is the ability to assemble, analyze, and interpret data from various sources (the so-called afferent side of the learning health system cycle). Thus, a fundamental requirement for an organization to succeed as a learning health system is its ability to efficiently and effectively translate evidence into practice. Unfortunately, translation of evidence-based medicine into clinical practice is accompanied by a considerable delay (~17 years), which has real consequences, both in terms of adverse health outcomes and overall healthcare costs. There are many reasons for this evidence-practice gap, but we hypothesize that a few key reasons for this delay include: 1) identifying populations of patients that may be eligible for novel diagnostics and/or therapeutics; 2) measuring variation in care delivery; and 3) measuring the impact of clinical decisions and interventions on health outcomes, healthcare utilization, as well as patient reported outcomes.

SmartRx is a natural language processing (NLP) based software platform that enables automated, real-time, querying of the entire electronic health record for specific patient cohorts with targeted inclusion and exclusion criteria. The key value proposition for the product includes the ability to: a) measure gaps in quality of care at an individual and population level and thus, close the evidence-practice gap; and b) leverage existing practice-based medicine (historic data) to drive clinical decision making and evidence generation, in cases where high quality evidence is lacking. Beyond enabling the learning health system, there are many other examples that could demonstrate the value of this tool, such as: a) therapeutic appropriateness as a measure of quality of care; b) clinical trial screening and recruitment; and c) prior authorization for medications. We aim to test the software platform in all the above-mentioned scenarios to demonstrate the performance and scalability of our solution.



**Figure 1.** Repositioning of approved medications (red) by reference to a panel of medications (orange) with desired drug burden.



**Figure 2.** Time to delirium by quartile of drug burden based predicted risk.

## Computable Pharmacology Through Drug Burden Scoring



### Thomas McCoy, MD

Director of Research, Center for Quantitative Health, MGH;  
Assistant Professor of Psychiatry & Medicine, HMS

[THMCCOY@PARTNERS.ORG](mailto:THMCCOY@PARTNERS.ORG)

Drawing on clinical expertise, we developed and validated composite drug burden scores for risk of outcomes of interest (falls and delirium, to date). These scores are not merely prediction tools; rather, they enable us to stratify side-effect risk within medication classes, enabling outcome-specific recommendations among clinically equivalent options, and integrate over full medication lists, thereby detecting significant risks to a patient that emerges from many small contributions. In short, this is a model that is inherently actionable.

This actionable nature arises from the approach taken. The medication approval and monitoring process generates extensive data on the desired and undesired effects of medications. Using these data to position medications in the space of human biological activity enables comparisons between and aggregation across medications without reference to a specific clinical cohort. Further, because effects captured in approvals and monitoring have known clinical significance, this feature space can be directly reasoned about by clinical experts.

To develop this widely applicable approach to computing over prescription medications we leveraged the public pre- and post-market data from the medication approval process conceptualized as an *in vivo* bioassay of medication effects. Unlike *in silico* trials of people on a particular medication, we directly parameterize the medications themselves thereby reducing the complexity of clinical prediction and expanding the range of applications beyond the initial clinical prediction validation.

Building from the predictive validation of the approach, we have enhanced aggregated risk scoring with the incorporation of CYP450 interaction models resulting in natural integrations with commercial CYP genetic testing. Working with pre-clinical partners, we have shown this approach enables repositioning of approved medications as candidate treatments. Finally, we are refining semi-supervised approaches to the time intensive step of capturing expert clinical knowledge and expanding our source data beyond the formal regulatory data to include large consumer datasets as a complementary assay of medication effect.

## A Machine Learning Algorithm to Reduce Costs and Increase Quality

---



### Ziad Obermeyer, MD

Assistant Professor, Emergency Medicine, BWH; Assistant Professor, HMS

[zobermeyer@bwh.harvard.edu](mailto:zobermeyer@bwh.harvard.edu)

Diagnostic tests are the lynchpin of medical care. From a health point of view, bad testing can lead to missed or incorrect diagnoses. From a cost point of view, bad testing is a major source of unneeded expenditures—the majority of the care labeled as “low value” by the Choosing Wisely campaign were diagnostic tests.

And yet testing decisions are notoriously hard – especially in a world where every patient is a “big data” challenge. Electronic health records hold exponentially increasing amounts of data as patients age, medical complexity increases, and technology expands rapidly. Doctors must absorb all this information in deciding which, if any, tests to deploy.

My team is building algorithms to aid physicians in these testing decisions – we have one completed prototype and several in the works. Importantly we are finding that these algorithms add value in two different ways. First, they can save costs – by dramatically reducing low-value testing of patients in whom nothing is found. Second, they can actually improve health outcomes – by identifying large pools of “misses”: untested patients who actually could have been caught, had they been tested in time. I will discuss what we might need to take these prototypes to scale as well as the kind of partnerships needed to build and apply these algorithms.

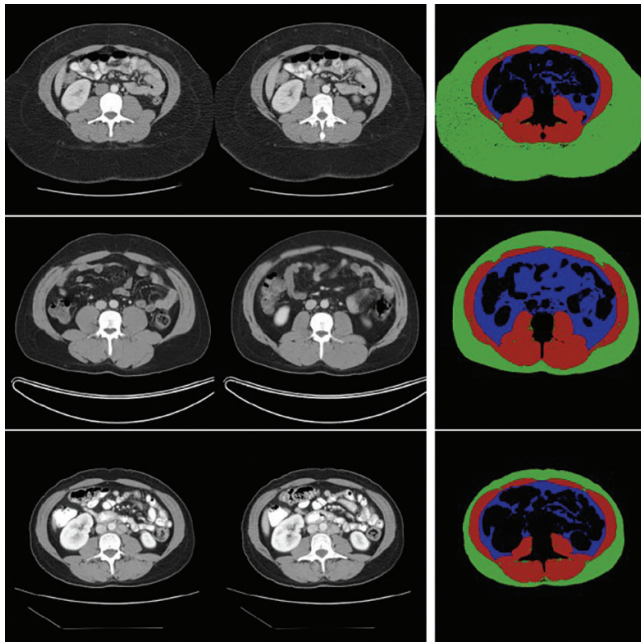


Figure 1. Human and AI selections and AI segmentations.

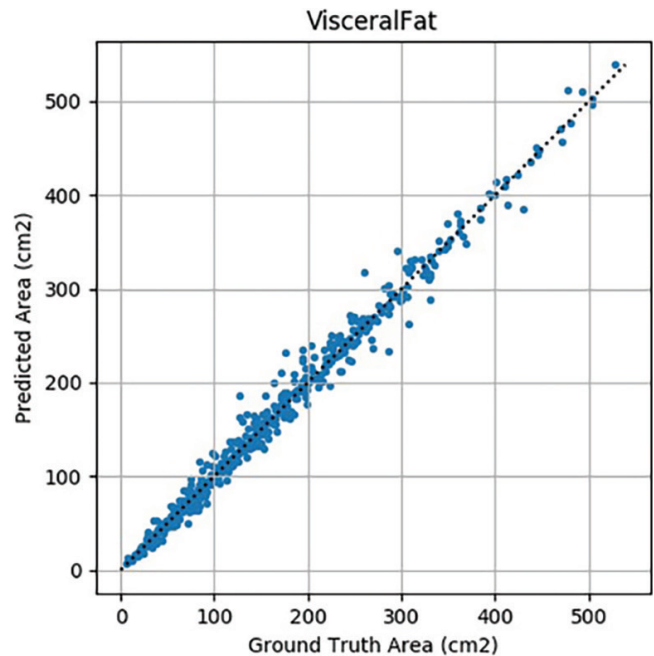


Figure 2. Comparison of manual to AI measurements of visceral fat.

## Extracting Muscle and Fat Metrics from Clinical CT Images Using Neural Networks



### Michael Rosenthal, MD, PhD

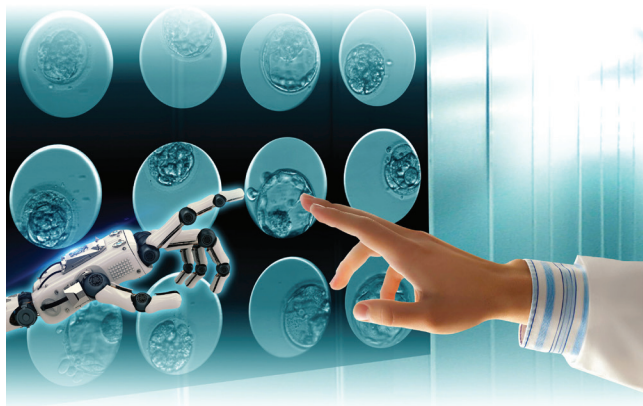
Radiologist, BWH; Assistant Professor, Radiology, HMS  
[mrosenthal@bwh.harvard.edu](mailto:mrosenthal@bwh.harvard.edu)

The amounts of muscle and fat in a person's body, known as body composition, are correlated with cancer risks, cancer survival, and cardiovascular risk. The current gold standard for measuring body composition requires time-consuming manual image selection and segmentation of CT images by an expert reader and has limited this research to well-funded investigators. This barrier to entry has prevented widespread use of body composition data in research and clinical care.

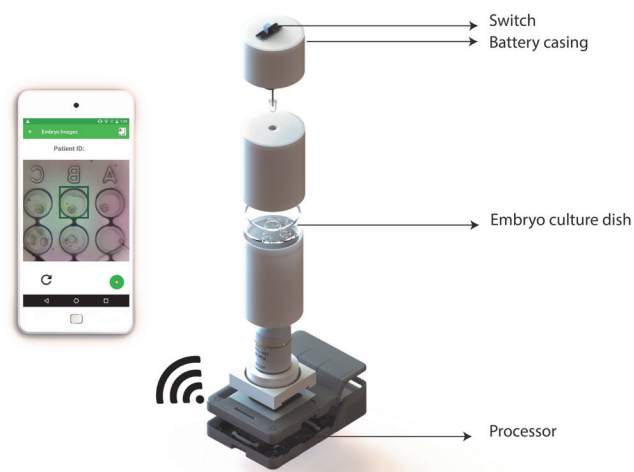
In this work, we describe a two-step process to fully automate the analysis of CT body composition using two distinct neural network models to select the CT slice and to perform segmentation. We train and test our methods on independent cohorts. Our results show Dice scores (0.95-0.98; a measure of agreement) and correlation coefficients ( $R=0.99$ ) that are favorable compared to human readers. The analysis takes 1 second per study on commercial GPU hardware.

These results suggest that fully automated body composition analysis is feasible and could be incorporated into routine clinical imaging with no negative impact on radiologist workflow.





**Figure 1.** AI-empowered computing algorithm for embryo assessment.



**Figure 2.** AI-empowered, low-cost, portable optical system for embryo data acquisition, analysis, and reporting through a cellphone wireless connection.

## AI-Empowered Optical System for In-Vitro Fertilization



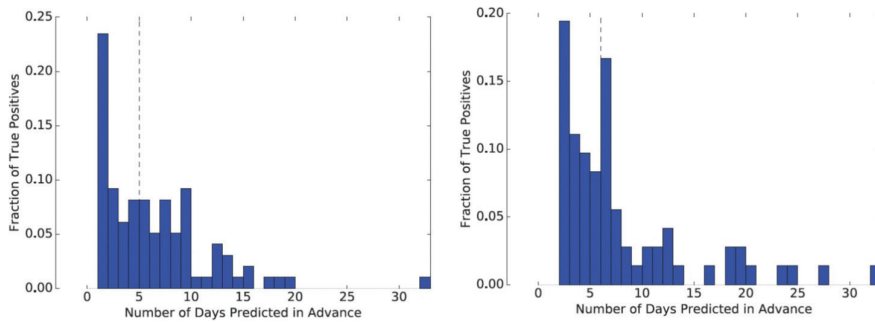
### Hadi Shafiee, PhD

Director, Laboratory of Micro/Nanomedicine and Digital Health, BWH; Assistant Professor, HMS

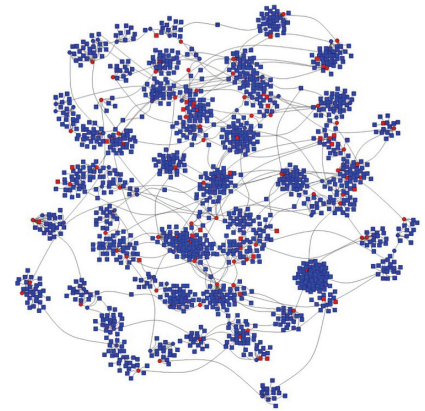
[hshafiee@bwh.harvard.edu](mailto:hshafiee@bwh.harvard.edu)

Infertility is an underestimated healthcare problem that affects over 48 million couples globally and is a cause of distress, depression and discrimination. The human fertility market is projected to be around \$40B globally by 2020. Although assisted reproductive technologies (ART) such as in-vitro fertilization (IVF) has alleviated the disease burden to an extent, it has been inefficient. According to center for disease control's (CDC's) 2015 Fertility Clinic Success Rates Report, there were 231,936 IVF cycles performed in the United States in 2015, out of which, only 60,778 led to live births, i.e. 26% success rate. Therefore, 74% of patients had to go through multiple IVF cycles. IVF remains as an expensive solution costing \$7000 - \$20,000 per ART cycle in the US that may not be covered by insurance with many patients requiring multiple cycles to achieve pregnancy. Therefore, the cumulative cost for only one additional IVF procedure for all of those with unsuccessful IVF would have been more than \$1 B in the US alone.

The poor efficiency of the IVF process enforces the majority of the patients to go through several cycles of treatment to obtain a successful birth. The major issue is related to the inefficient and subjective process of identifying generated embryos with the optimum possibility of producing a child. Therefore, currently, the IVF clinics in many countries transfer more than one embryo during each cycle of treatment to maintain the pregnancy rate at an acceptable level; however, this process may lead to multiple pregnancies, which has negative consequences including maternal haemorrhage, pre-eclampsia, uterine rupture, operative delivery, etc. The selection of the best quality embryo during an in-vitro embryo culture and before its transfer to a patient remains the most important factor in achieving successful ART-based pregnancies and still is a significant challenge in the field of embryology. We overcome this challenge by employing an artificial intelligence (AI)-empowered algorithm pre-trained with 1.4 million images to automatically classify human embryos at three discrete clinically relevant stages of embryonic development.



**Figure 1.** Measuring how far in advance the model correctly identifies cases.



**Figure 2.** Network representation of *C. diff* transmission pathways extracted from MGH data.

## Prediction as Prevention: A Data-Driven Approach to Identify Patients at Risk for Infection



### Erica Shenoy, MD, PhD

Associate Chief, Infection Control Unit, MGH; Assistant Professor, HMS  
[eshenoy@mgh.harvard.edu](mailto:eshenoy@mgh.harvard.edu)

*Clostridium difficile* (*C. diff*) is responsible for approximately 500,000 infections per year, of which 66% are healthcare-associated. Considered one of the most urgent microbial threats by the Centers for Disease Control and Prevention, estimates of the excess costs of *C. diff* to the healthcare system range from \$897 million to over \$4 billion.

Early identification of patients at risk for *C. diff* could: 1) improve patient outcomes through earlier treatment, 2) decrease transmission through early institution of infection control measures, and 3) reduce healthcare expenditures. Many previous efforts have focused on using a small number of variables to build predictive models, however, recent work has demonstrated that more effective risk stratification models can be constructed using machine learning techniques applied to large sets of covariates (features) drawn directly from the electronic health record. These models leverage rich temporal signals that arise from high-dimensional time-series data, boosting discriminative power.

Our team at MGH, MIT, and the University of Michigan has validated a flexible approach to building risk prediction models for CDI at two separate institutions, including over 250,000 patients. We extracted patient demographics, admission details, patient history, and daily hospitalization details, resulting in thousands of features, and then used these features to build predictive models. The UM and MGH models achieved area under the receiver operating characteristic curve (AUROC) values of 0.82 and 0.75, respectively. On average, the models were able to identify patients 5 days earlier than clinical diagnosis (Fig. 1).

Our next step is to evaluate the efficacy of using data-driven risk models to guide interventions that can impact clinical and economic outcomes. These models can also be used in clinical trial design for targeting patient recruitment, optimizing clinical trial efficiency. Finally, we are exploring approaches to characterize how *C. diff* spreads among hospitalized patients (Fig. 2), which would identify opportunities to interrupt transmission events.

## Artificially Intelligent Sleep Analysis with Deep Neural Networks



### Brandon Westover, MD, PhD

Director, MGH Clinical Data Animation Center, MGH; Assistant Professor, HMS

[mwestover@mgh.harvard.edu](mailto:mwestover@mgh.harvard.edu)

Sleep disorders affect 20% of adults and are significant risk factors for cardiometabolic and neurodegenerative diseases, impaired performance, and decreased quality of life<sup>1-7</sup>. Medical analysis of sleep data is a manual task of visually annotating three major event types: sleep stages, sleep disordered breathing, and limb movements. Attempts to automate sleep analysis have been hampered by the complexity of polysomnographic (PSG) signals and physiological heterogeneity between patients. Deep neural networks have recently achieved expert-level performance in interpreting complex medical data. Here we demonstrate automated scoring of overnight PSG recordings using a combination of deep recurrent and convolutional neural networks (RCNN) trained on physiological signals and clinical labels designating sleep stages, sleep apnea events, and limb movements from 10,000 clinical PSGs – two orders of magnitude larger than previous datasets. The RCNN reproduces PSG diagnostic results for sleep staging, apnea detection and limb movement detection with accuracy of 87.56%, 88.24% and 84.72%, respectively, a level of performance comparable to human experts. The resulting RCNN model performs equally well on a large (>5000) independent set of research trial PSGs. Our deep neural network AI technology is poised to meet a rapidly rising demand for low-cost home sleep monitoring devices. This technology will not only extend the reach of clinical sleep experts far beyond specialty clinics, but will also address the larger market of consumers seeking scientifically valid tools to help monitor and optimize brain health.



# Mobile Apps: Bridging the Mental Health Treatment Gap

## Sabine Wilhelm, PhD

Chief of Psychology; Director, OCD and Related Disorders Program, MGH; Professor, HMS

[swilhelm@mg.harvard.edu](mailto:swilhelm@mg.harvard.edu)

Barriers to accessing evidence-based mental health treatment include a lack of available providers, high costs, and stigma. Technology-delivered interventions, such as Smartphone apps, present a promising avenue for reducing barriers. To this end, in partnership with Telefónica Alpha, the Massachusetts General Hospital (MGH) OCD and Related Disorders program is developing and pilot testing a cognitive behavioral therapy (CBT) app for individuals with body dysmorphic disorder (BDD).

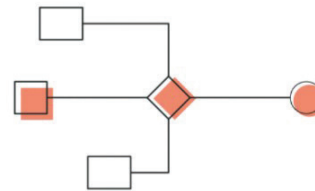
This collaboration capitalizes on the synthesis of MGH's clinical and research expertise, along with Telefónica Alpha's design and technology expertise, to produce a high quality, evidence-based, mobile CBT app for a severe psychiatric body image disorder. To date, we have developed a beta version of the app and are testing its feasibility, acceptability, and preliminary efficacy in an iterative open pilot trial. After establishing a finalized version, we will further test the app's efficacy in a randomized waitlist-controlled trial.

Next steps: We are collecting passive sensor data (e.g., mobility, sleep) that will be used to personalize treatment in later app versions. For example, we aim to passively detect clinically relevant changes to users' mental health status (e.g., increased depression or BDD symptoms, times when users get "stuck" doing rituals), to provide real-time intervention suggestions delivered via the app. Moreover, CBT is the first-line psychosocial treatment for many psychiatric disorders, beyond BDD. We therefore designed the app to incorporate each of the primary universal components of CBT and we plan to adapt our CBT app for other clinical use cases, by making disorder-relevant modifications.

## How does it work?

As you go through these sections, you'll start learning new skills, like how to spot unhelpful thoughts.

You'll choose goals for the kind of life you want to live. Then, later, you'll do exercises to help you achieve those goals, one step at a time.

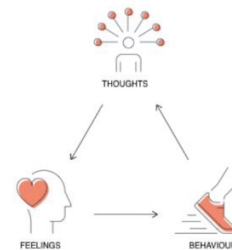


## SPOTTING UNHELPFUL THOUGHTS



## How thoughts and feelings are connected

Our thoughts play a huge part in how we feel and react to things.

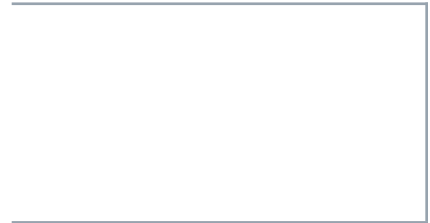


Psychoeducation is a hallmark feature of the app.



# WORLD MEDICAL INNOVATION FORUM sponsors

The 2018 World Medical Innovation Forum was made possible through the generosity of the following sponsors.



## stakeholder



## strategic

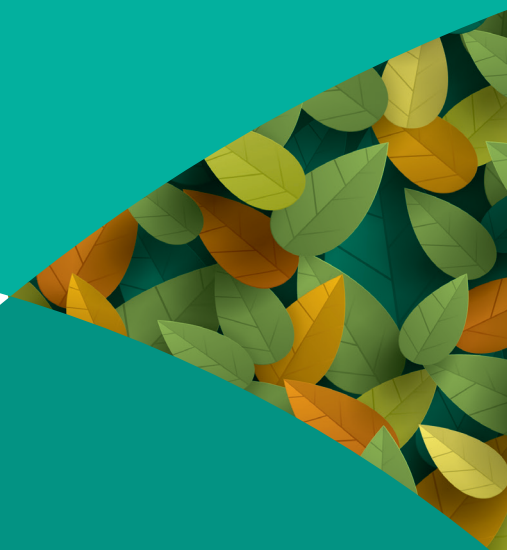


## collaborator

## media



**Be a part of the future.**



# Partners HealthCare Biobank Disease Challenge

**April 23 - June 29**

Online Registration

**July 2 - August 19**

Practice on Platform

**September 12 -  
October 10**

Competition

---

---

## Prize Awards

First - \$15,000

Second - \$10,000

Third - \$5,000

## About Our Challenge

Partners HealthCare is hosting a “Biobank Disease Challenge,” an artificial intelligence and machine learning data analytics competition and it is open to researchers across the United States. The goal of this competition is to enable major translational data science players to leverage the Partners HealthCare Biobank in order to develop better phenotypic algorithms for clinical and basic research.

Registration will be open on April 23 and close on June 29, 2018 to 50 teams. Prize money will be awarded to the top 3 teams.

To learn more and get up to date information,  
visit us at <http://www.partners.org/biobankchallenge>



FOUNDED BY BRIGHAM AND WOMEN'S HOSPITAL  
AND MASSACHUSETTS GENERAL HOSPITAL



Partners HealthCare

# INNOVATION

This World Forum is brought to you by Partners HealthCare Innovation, the nearly 120 person business development unit responsible for the worldwide commercial application of the capabilities and discoveries of Partners' 74,000 employees. Innovation supports the research requirements of its 3,600 Harvard Medical School faculty and research hospitals. It has responsibility for industry collaborations, venture investing, international consulting, licensing, innovation management, company creation, technology marketing, open innovation alliances, and workforce development.

## DEPARTMENTS

- Boston Biomedical Innovation Center
- Business Development
- Innovation Operations
- Licensing and Innovation Management
- Market Sectors
- Open Innovation – Alliances and Fellows
- Partners HealthCare International
- Partners Innovation Fund

## FUNCTIONS

- Business development
- Company creation
- Industry collaborations
- Innovation management
- International consulting
- Licensing
- Research translation strategy and funding
- Technology marketing
- Venture investing
- Workforce capacity building for commercial collaborations



**PARTNERS**<sup>®</sup> | INNOVATION  
HEALTHCARE

215 First Street, Suite 500, Cambridge, MA 02142  
857-307-2400 | [innovation.partners.org](http://innovation.partners.org)

**PARTNERS**  
HEALTHCARE | INNOVATION

215 First Street, Suite 500 | Cambridge, MA 02142  
857-307-2400 | [innovation.partners.org](http://innovation.partners.org)

**PARTNERS**  
HEALTHCARE | FOUNDED BY BRIGHAM AND WOMEN'S HOSPITAL  
AND MASSACHUSETTS GENERAL HOSPITAL

